The Visual Representation of Facial Expression

Nichola Burton
Bachelor of Arts (1st Class Hons)

School of Psychology
The University of Western Australia

This thesis is presented for the degree of Doctor of Philosophy of The University of Western Australia

2015
Abstract

In this thesis, we examined the visual representation of facial expressions, and the mechanisms that underlie expression perception. We conducted three sets of studies of expression perception in an adult sample, each investigating a different aspect of the visual representation of facial expressions.

In Chapter 2 we compared two potential models of the neural coding of expressions. Expressions are typically modeled as being represented in a two-pool opponent system, relative to an implicitly-represented norm. However, the evidence for this model cannot rule out an alternative model in which one neural channel codes the center of the dimension, explicitly coding the norm, in addition to one or more channels coding each end of the dimension. This type of multichannel coding system has been shown to underlie the coding of head orientation and gaze direction. We used an adaptation paradigm capable of discriminating between these two models, and found evidence that supports the two-pool opponent-coding model. Additionally, after adapting to the center of an expression morph trajectory, participants were more likely to identify expressions on that trajectory as resembling the endpoint expressions. This finding may reflect increased sensitivity following adaptation, and is the first evidence of a potential functional benefit of expression adaptation. These results have recently been published in the Journal of Vision.

In Chapter 3 we aimed to determine which expression functions as the norm of expression-space. There is evidence that expressions are represented in a multi-dimensional “space,” relative to a central, norm expression. However, there is no consensus on what this norm is likely to be. Two contenders proposed in the literature are the neutral expression and the average expression. In this chapter compared these two potential norms using an adaptation paradigm. We found no
difference between the two potential ‘norms’, a finding that suggests that they may be equally good approximations of the true norm.

In Chapter 4 we examined the timecourse of the expression aftereffect. Experiment 1 investigated how adaptation duration and test duration affected aftereffect strength. The expression aftereffect followed the same classic timecourse pattern that has been found for lower-level aftereffects, and for identity and figural face aftereffects. This timecourse pattern is evidence that the expression aftereffect has a perceptual locus, and is not simply a result of post-perceptual response biases. Experiment 2 explored the longevity of the expression aftereffect. Relatively brief adaptation (8 sec) induced an aftereffect that lasted at least 32 seconds (our longest test duration), a surprising finding given the dynamic nature of facial expressions. This persistence may allow the visual system to integrate expression information over time, helping to keep limited perceptual resources calibrated to the most useful range of variation. These studies are currently under revision for publication in the Journal of Vision.

The current studies provide information about the way that facial expressions are visually represented. These findings show that expressions are coded in a two-pool opponent-coding system, relative to an implicitly-represented norm. We show that expression aftereffects are perceptual in nature, confirming that this coding system adapts with exposure to faces. We also found that these adaptation aftereffects last for long enough to have an effect on day-to-day social interactions, and found evidence that adaptation plays a functional role in expression perception. We were unable to clearly determine whether the average or neutral expression is a better approximation of the norm against which expressions are coded. In Chapter 5 we discuss the relationship between the representation of facial expression and facial identity.
Contents

Abstract ........................................................................................................... ii

Contents .......................................................................................................... iv

Acknowledgements ........................................................................................ viii

Manuscripts for Publication ........................................................................... ix

Chapter One: General Introduction ................................................................. 1

Themes in the facial expression literature ....................................................... 2

The scope of the present thesis ....................................................................... 3

Summary: the aims of this thesis in the context of the current state of the literature ......................................................................................................................... 12

The present thesis: studies ............................................................................... 13

References ....................................................................................................... 17

Chapter Two: How is Facial Expression Coded? ............................................. 27

Abstract .......................................................................................................... 28

Introduction ...................................................................................................... 29

Method ............................................................................................................. 36

Participants ..................................................................................................... 36

Stimuli .............................................................................................................. 36

Procedure ....................................................................................................... 38

Results ............................................................................................................. 41

Discussion ...................................................................................................... 49

Acknowledgments ........................................................................................... 57

References ....................................................................................................... 58

Supplementary Material: The Mixed Multinomial Logit Model ............... 63

References ....................................................................................................... 70
Chapter Three: Is a Neutral or an Average Expression Closer to the Norm of Expression Space? ................................................................. 72

Abstract............................................................................................................. 73
Introduction........................................................................................................ 74
Method............................................................................................................. 78
  Participants...................................................................................................... 78
  Stimuli........................................................................................................... 79
  Procedure....................................................................................................... 84
    Training....................................................................................................... 84
    Baseline Block........................................................................................... 85
    Adaptation Block...................................................................................... 86
Results............................................................................................................. 87
Discussion....................................................................................................... 91
Acknowledgements......................................................................................... 94
References..................................................................................................... 95
Supplementary Materials................................................................................ 99
  A: Matching adaptor similarity................................................................. 99
  B: The mixed logit model......................................................................... 102
References..................................................................................................... 109

Chapter Four: The Timecourse of Expression Aftereffects....................... 111

Abstract.......................................................................................................... 112
Introduction..................................................................................................... 113
Experiment 1.................................................................................................. 116
  Method......................................................................................................... 117
  Participants.................................................................................................. 117
  Stimuli.......................................................................................................... 117
Acknowledgements

I am grateful to many people who have helped to make this thesis happen.

First and foremost, I thank my supervisors Gill Rhodes and Linda Jeffery, who have been an endless source of good humour, enthusiasm and excellent advice. I am so lucky to have stumbled into your lab, and I’m so glad I have been able to stay.

Thanks as well to the rest of the members of the FaceLab, who make the lab such a warm, welcoming environment. There’s a reason no one wants to leave! Particular thanks to Libby for taking care of all of us, and to Yong, Clare and Bianca for pushing me through the last few months.

Thanks to Chris and Andy, my generous and welcoming collaborators in Bristol.

Thanks to Mum, Thomas and James, and to Max and Ollie, for your support and friendship.

Thanks to Dad, for all the cross-disciplinary discussions, and the chocolate.

And finally thanks to Jo, who always knows how to make me laugh.

The work presented in this thesis was supported by an Australian Postgraduate award and a University of Western Australia Completion scholarship.
Manuscripts for Publication

This thesis is submitted as a series of manuscripts that have been prepared for publication in international journals, each of which has been co-authored. The bibliographical details of each manuscript and where it appears in the thesis are outlined below:


III. Burton, N., Jeffery, L., Bonner, J., & Rhodes, G. The timecourse of expression aftereffects. Journal of Vision, under revision for publication. (Chapter 4)

For each manuscript, I (Nichola Burton) was the main contributor, responsible for the majority of experimental design, data collection, statistical analyses and writing. For all three manuscripts, Linda Jeffery and Gillian Rhodes also contributed to the experimental design, statistical analysis and writing. The contributions of our other co-authors are listed below:
I. Andy Calder: contributed to experimental design.

II. Andy Skinner: provided stimuli, contributed to experimental design, analyzed data for preliminary adaptor matching, contributed to interpretation of results and writing.

Chris Benton: provided stimuli, contributed to experimental design, contributed to data analysis, interpretation of results and writing.

III. Jack Bonner: assisted with stimulus preparation and data collection for Experiment 2.

Each author has given permission for all work to be included in this thesis.

Nichola Burton (candidate)  Date

Gillian Rhodes (coordinating supervisor)  Date
Chapter One

General Introduction
General Introduction

Faces are important social stimuli that we encounter frequently throughout our daily experiences. We are able to quickly and easily extract many kinds of information about a person from their face, including their identity, race, age and sex. One particularly important kind of information that we extract from a face is expression. Facial expressions can reflect internal states, or can be used as intentional signals for communication. Our ability to read expressions can therefore help us to predict the intentions and behavior of others, and to modulate our own behavior in response.

Themes in the facial expression literature

The importance of facial expressions as social stimuli means that they have been the subject of extensive research. One major theme has been the connection between an expression and the internal state of the person producing that expression. A fundamental question for this research is whether certain emotional states or circumstances reliably evoke particular expressions (Darwin, Ekman, & Prodger, 1998; Ekman, 1993). An associated question concerns whether expressions are learned social signals, or automatic, inbuilt responses to internal states (Elfenbein & Ambady, 2002). Ekman and his colleagues have championed the universality of facial expressions, finding that a set of “basic” emotional expressions is widely recognized across cultures (Ekman, 1970, 1971). A program of research has therefore been focused on identifying the particular muscle movements that signal these expressions (Ekman, Friesen, & Hager, 1978).

Attention has also been paid to the way that meaning is read from facial expressions. For instance, there has been debate about whether expressions are interpreted in a dimensional framework – for instance, in terms of their positions
on valence and arousal dimensions (Russell, 1980; Russell & Bullock, 1985; Schlosberg, 1952, 1954) - or as members of discrete emotion categories (Young et al., 1997). Research has also investigated the role played by context and expectations in the interpretation of expressions (Aviezer et al., 2008), and the way that expressions capture and direct attention (O'Toole, DeCicco, Hong, & Dennis, 2011).

**The scope of the present thesis**

One area that has been relatively under-studied is the process that occurs between the production and interpretation of facial expressions: namely, the processing of facial expressions as visual stimuli. The way that expressions are visually processed and represented is integral to our ability to use these stimuli as a source of social information. This thesis aims to broaden our understanding of this important aspect of the psychology of facial expressions. Although the visual representation of facial expressions has not been widely investigated, we know more about the visual representation of facial identity. In this thesis, we set out to determine what mechanisms underlie the perception of facial expression, using what has been discovered in facial identity research to guide our investigation. In the sections below we outline what we currently know about how facial expression is represented, and what we might be able to learn from facial identity research.

*Adaptability in perception of facial expression and facial identity*

One feature of the visual representation of facial expressions that may benefit expression perception is that this representation appears to be adaptable. Many aspects of visual perception show adaptation, a phenomenon in which the appearance of a stimulus is affected by what has been seen before (Clifford & Rhodes, 2005). The classic example is motion adaptation: After watching the
downward motion of a waterfall, subsequent stationary stimuli appear to move upwards (Anstis, Verstraten, & Mather, 1998). The perceptual bias created by adaptation is known as an aftereffect. These aftereffects can be extremely useful for perceptual research, because they reflect changes in the neural mechanisms that underlie perception. For instance, the motion aftereffect corresponds to reduced responsiveness in the neurons that respond preferentially to the adapted direction of motion (Van Wezel & Britten, 2002).

Adaptation may be a functional process that aids visual perception. Our visual systems must use limited resources to process a wide range of visual inputs. Adaptation can help to calibrate perception by shifting our responsive range to the currently encountered range of stimulus values (Barlow, 1972; Clifford, Wenderoth, & Spehar, 2000). Evidence for improved discrimination following adaptation has been found for many low level visual properties, including motion (Phinney, Bowd, & Patterson, 1997; but see Hol & Treue, 2001), speed (Clifford & Langley, 1996; Krekelberg, Van Wezel, & Albright, 2006), orientation (Clifford, Wyatt, Arnold, Smith, & Wenderoth, 2001; Regan & Beverley, 1985; but see Barlow, MacLeod, & van Meeteren, 1976), and contrast (Abbonizio, Langley, & Clifford, 2002; but see Barlow et al., 1976; Maatenen & Koenderink, 1991).

Adaptation aftereffects have been demonstrated for faces, as well as for lower-level visual stimuli like the ones described above. For instance, adapting to a particular identity will make subsequent faces look less like that identity (Leopold, O’Toole, Vetter, & Blanz, 2001). These aftereffects survive changes in the size, colour, contrast and orientation of a face (Rhodes et al., 2005; Watson & Clifford, 2003; Yamashita, Hardy, De Valois, & Webster, 2005; Zhao & Chubb, 2001), suggesting that they tap higher-level visual encoding of faces. As in low-
level visual adaptation, there is evidence for a functional benefit of adaptation in face perception. There is a positive relationship between the adaptability of face perception and performance in identity recognition tasks (Dennett, McKone, Edwards, & Susilo, 2012; Rhodes, Jeffery, Taylor, Hayward, & Ewing, 2014; Rhodes et al., 2015). In addition, identity discrimination thresholds are lowest around the average, which should lie close to the current adapted state (Wilson, Loffler, & Wilkinson, 2002; but see Rhodes, Maloney, Turner & Ewing, 2007). Rhodes, Watson, Jeffery, and Clifford (2010) have also found that adapting to an average Asian or Caucasian face improves identification within the adapted race. These findings suggest that adaptation may also have a functional role for high-level face stimuli like identity.

Facial expression perception is also adaptable: adapting to an expression will bias subsequent perception away from that expression. Several studies have demonstrated this effect using expression trajectories made by morphing from one expression to another (e.g. anger-fear, happy-sad). Adapting to the expression at one end of the trajectory biases perception towards the expression at the other end of the trajectory (Benton, 2009; Bestelmeyer, Jones, DeBruine, Little, & Welling, 2010; Ellamil, Susskind, & Anderson, 2008; Fox & Barton, 2007; Vida & Mondloch, 2009; Webster, Kaping, Mizokami, & Duhamel, 2004). Similarly, adaptation makes faces on a trajectory running from the adaptor to a neutral face appear less like the adapted expression (Adams, Gray, Garner, & Graf, 2010; Hsu & Young, 2004; Pell & Richards, 2011).

We can find out more about the way in which facial expressions are represented by identifying the direction in which perception is biased following adaptation. The above studies demonstrate that adaptation biases perception away from the adaptor. We can determine more specifically whether perception is
biased selectively in one direction or simply away from the adaptor in all directions by adapting participants to anti-expressions. An anti-expression is the visual opposite of an expression, made relative to an average expression or some similar reference point. For instance, where fear has higher eyebrows than the average expression, anti-fear would have lower eyebrows, and so on (see Figure 1). Adapting to an anti-expression biases perception selectively towards the opposite expression (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Cook, Matei, & Johnston, 2011; Juricevic & Webster, 2012; Skinner & Benton, 2010, 2012a). Anti-expressions are ideal adaptors for experimental designs that make predictions about the direction of the aftereffects that result from adaptation (e.g. Cook et al., 2011). These stimuli are also useful adaptors because adaptation to an anti-expression biases perception towards an expression that is familiar to and easily named by participants.

![Fear (left) and anti-fear (right). The feature positions of these two expressions are opposite one another relative to the average expression (center).](image)

*Figure 1:* Fear (left) and anti-fear (right). The feature positions of these two expressions are opposite one another relative to the average expression (center).

Expression aftereffects survive changes in stimulus size and free eye movements (Burton et al., 2013; Skinner & Benton, 2010, 2012a), and are still produced (although smaller in size) if adaptor and test show different images of the same person, or even different people (Fox & Barton, 2007; Skinner &
Benton, 2012a; Vida & Mondloch, 2009). These findings indicate that expression aftereffects are not simply a result of retinotopic low-level adaptation. Nor can these aftereffects be explained by adaptation of later, cognitive representations of emotion. Fox and Barton (2007) demonstrate that expression aftereffects can be generated by images of facial expressions, but not by other visual, auditory or verbal stimuli representing the same emotions, indicating that expression adaptation occurs at the level of the face rather than the emotion.

Like identity adaptation, expression adaptation might help to focus expression perception around the current range of expressions that we experience at a given time, helping to distinguish between subtle differences. However, the benefits of adaptation for expression perception are less well established than the benefit for identity perception. Rhodes et al. (2015) found a correlation between adaptability of expression perception and expression recognition ability. To our knowledge this is the only study that has directly shown a functional benefit of adaptation for expression perception.

**Norm based coding of facial expression and facial identity**

Another mechanism of visual representation that may contribute to expression perception is norm-based coding. An important model of face coding is “face space” (Valentine, 1991), originally developed to describe the representation of facial identity (Figure 2). In this model, faces are represented in a multidimensional “space” defined by the dimensions on which faces are perceived to vary. What exactly these dimensions code is not known, but they are generally theorized to be dimensions derived from the natural statistics of faces – perhaps simple variation in feature positions, or more complex dimensions representing something like the principal components that can be extracted from image variation in faces (Robbins, McKone, & Edwards, 2007). The distance
between faces in face-space reflects how similar, and therefore how confusable, they are. The center of face space represents the central tendency of faces, and is the reference point against which the positions of other faces are coded (Rhodes & Leopold, 2011). Norm-based coding is efficient, because we do not need to encode the undiagnostic aspects of a face that are common to most faces. Rather, coding is focused on what is distinctive about a face (Rhodes & Leopold, 2011; Webster & MacLeod, 2011).

Figure 2: A simple two-dimensional schematic of face space. The space is defined by coding dimensions, which represent ways in which faces are perceived to vary. At the center of the space is the norm, which represents the central tendency of faces. Faces are represented by their position in the space relative to the norm.

Norm-based coding may be implemented by a two-pool opponent coding system (Rhodes & Jeffery, 2006; Robbins et al., 2007; Tsao & Freiwald, 2006). In this type of coding system, each dimension of face-space is coded by two pools of
neurons, one responding maximally to one end of the dimension and one responding maximally to the other end of the dimension (Figure 3A). The norm is represented implicitly as the point at which both pools respond equally. The strength of response created by a stimulus is proportional to how far that stimulus lies from the norm, so resources are focused on what is distinctive. This model is in line with neuroimaging in humans (Loffler, Yourganov, Wilkinson, & Wilson, 2005) and single-cell recordings in monkeys (Leopold, Bondar, & Giese, 2006), which both show increased face-related activation as faces move further from the average.

![Figure 3: A two-pool opponent coding model, showing neural responsiveness plotted along a coding dimension. A) Neural response curves before adaptation. Two neural pools code the dimension, one responding maximally to values at each end of the dimension. The norm is represented implicitly as the point where both pools respond equally. B) Neural response curves after adaptation to a stimulus from the right-hand end of the dimension. The responsiveness of the pool of neurons that preferentially responds to this end of the dimension is suppressed by adaptation. The norm (point of equal activation) shifts, such that subsequently-seen stimuli now appear more like the left-hand end of the dimension.](image)

In a two-pool opponent model, adaptation reduces the responsiveness of the neural pools that code the adaptor, altering the balance of responsiveness and
shifting the norm to the new point of equal activation (Figure 3B). For instance, we can imagine a coding dimension that represents nose width. Adapting to a wide nose reduces the responsiveness of the neurons that preferentially respond to wider noses, shifting the norm towards the wide end of the dimension. Noses that previously appeared wide now lie closer to the norm, and as a result these noses appear less wide than they did before.

The mere presence of face aftereffects is not sufficient to establish that face representation is norm based. Other types of coding system can also produce adaptation aftereffects – for instance, we see aftereffects for line orientation, which is coded in a multichannel system (Clifford, 2002). However, we can test for opponent, norm-based coding by examining the patterns of aftereffects that are produced by different adaptors. Important evidence for norm-based coding of facial identity comes from the near-far paradigm, which compares the aftereffects produced by adaptors that are either near to or far from the norm (Robbins et al., 2007). In a norm-based coding system, more extreme adaptors (adaptors that are further from the norm) should produce larger aftereffects than less extreme adaptors, because they induce stronger activation of the neurons that code that end of the dimension, and so create a larger suppression of those neurons. In contrast, in a multichannel system stimuli along a dimension are coded by a series of narrowly-tuned neural channels. The more extreme the adaptors are, the less likely it is that they will affect the responsiveness of neurons that code the test face. Therefore, in this type of coding system more extreme adaptors should in fact produce smaller aftereffects than less extreme adaptors. The strength of an identity adaptor can be adjusted by morphing between that particular identity and an approximation of the norm. More extreme identity adaptors produce larger aftereffects than less extreme adaptors, indicating that identity coding is norm-
based (Jeffery et al., 2011; McKone, Jeffery, Boeing, Clifford, & Rhodes, 2014, 2015; Robbins et al., 2007).

There is also some evidence that facial expression is represented in an opponent-coded system, with implicit representation of the norm. Adaptation biases perception towards the opposite of the adaptor relative to the norm, but does not bias perception along the orthogonal direction, suggesting a role for the norm in the coding of expression (Cook et al., 2011). In addition to this finding, most of the evidence for norm based coding of facial expressions comes from the near-far paradigm described above. More extreme adaptors have been shown to produce larger expression aftereffects than less extreme adaptors in several studies (Burton et al., 2013; Skinner & Benton, 2010, 2012a).

However, the near-far paradigm cannot provide conclusive evidence that expression is opponent-coded. Two other kinds of face-related information, head orientation and gaze direction, are coded by an alternative system, in which there are broadly-tuned channels coding the ends of the dimension, but also a channel which codes the center of the dimension (Figure 4). This type of coding may underlie the representation of other face-related information, such as gaze direction (Calder, Jenkins, Cassel, & Clifford, 2008) and head orientation (Lawson, Clifford, & Calder, 2011). This model produces the same predictions in the near-far paradigm as opponent coding, so our current evidence for opponent, norm-based coding of facial expression cannot rule out this alternative model. The presence of this central channel, which would respond maximally to the norm, means that this type of coding would lack the efficiency that has been put forward as a major benefit of opponent coding. So far, there has been no attempt to test whether an additional, central channel that explicitly codes the norm is used in the representation of facial expression.
Figure 4: The central-channel coding model, showing neural responsiveness plotted along a coding dimension. As well as the two pools found in the opponent coding model (Figure 3), this model has a third channel that is tuned to the center of the dimension, explicitly coding the norm.

Summary: The aims of this thesis in the context of the current state of the literature

The present thesis aims to investigate how facial expression is visually represented. We have identified findings from the facial identity literature that may be relevant to understanding the representation of facial expression. Two important aspects of the representation of facial identity that may benefit identity perception are 1) that facial identity is visually represented relative to a norm in an efficient opponent-coded system, and 2) that the visual representation of facial identity is adaptable, and this adaptability may have functional benefits for identity perception. There is some evidence that like identity, facial expression is opponent-coded (Burton et al., 2013; Skinner & Benton, 2010, 2012a), but this evidence cannot rule out the existence of an alternative model with explicit representation of the norm (Calder et al., 2008; Lawson et al., 2011). Representations of facial expression also appear to be adaptable (e.g. Hsu & Young, 2004; Skinner & Benton, 2010; Webster et al., 2004), but only one study offers evidence for a functional benefit for this adaptability in expression (Rhodes
et al., 2015). In the experiments described below we aim to address these gaps in the literature.

We also address a connected question about the visual representation of facial expressions: If expression is represented relative to a norm, then what is this norm expression? Finally, we consider a wider point: do expression aftereffects actually tap perceptual processes? The majority of the studies on the visual representation of facial expressions outlined above use adaptation aftereffects in their methods. It is therefore of vital importance to establish that expression aftereffects are perceptual in locus, not only to validate the methods used in the work presented here, but for the interpretation of much of the research that has come before.

**The present thesis: Studies**

In Chapter 2, we tested whether expression is opponent-coded, as currently suggested, or whether there is an additional central channel that codes the center of the coding dimension (Calder et al., 2008; Lawson et al., 2011). We used a paradigm that has not previously been applied to expression perception to distinguish between these two possible models, and found that expression is more likely to be coded by an opponent system than a central-channel system. Additionally, this paradigm allowed us to investigate the effect of adaptation on participants’ sensitivity to facial expressions. We found that adapting to an average expression increased participants’ ability to identify expressions on either side of that average, suggesting a possible functional benefit of adaptation for expression perception.

Our support for an opponent coding model of expression representation reinforces the importance of the norm in expression coding. However, there is disagreement in the literature on what this norm might be. In Chapter 2 we used
an average expression (created by morphing together images of the six basic emotional expressions, plus neutral) to approximate the norm. This type of “average” norm has been used in several studies (Abelson & Sermat, 1962; Burton et al., 2013; Cook et al., 2011; Skinner & Benton, 2010, 2012a, 2012b), and resembles the average identity used to approximate the norm in facial identity research (e.g. Jeffery et al., 2011; Leopold et al., 2001; Nishimura, Maurer, Jeffery, Pellicano, & Rhodes, 2008; Rhodes, Evangelista, & Jeffery, 2009; Rhodes & Jeffery, 2006). The efficiency of norm-based coding lies in the fact that the norm represents what is common, so that coding is focused on what is distinctive (Rhodes & Leopold, 2011). An average expression fulfills this role of representing what is common to most expressions. However, other researchers have suggested that the norm against which expressions are visually represented may actually be a neutral expression (Juricevic & Webster, 2012; Rutherford, Chattha, & Krysko, 2008; Rutherford, Troubridge, & Walsh, 2011). The neutral expression has the appearance of being “unexpressive”, which may make it analogous to white, the norm of colour space (Juricevic & Webster, 2012). In Chapter 3, we attempted to determine which of these two expressions is a better approximation of the true norm against which expressions are coded. We found no difference between the two potential norms using an aftereffect paradigm, and argue that both expressions may in fact be a reasonable approximation of the true norm, and can be used to approximate the norm in expression adaptation research.

In Appendix 1 we describe an initial attempt to identify whether the average or neutral expression is a better approximation of the norm. This experiment follows the same logic as the adaptation experiment described in Chapter 3, but uses a different method of measuring the size of the aftereffect (measuring changes in responses to a single test stimulus, rather than shifts in the
psychometric functions). We found that this method, although often used in aftereffect research, was not appropriate for comparing aftereffects between different adapt-test trajectories. We discuss in depth the circumstances in which this method is and is not appropriate, which we hope will prove useful for the design of future adaptation research.

Expression aftereffects indicate that expression perception is flexible, changing with experience. Experiments that use adaptation aftereffects to investigate the visual representation of facial expressions assume that this flexibility occurs at the visual level. However, it is also possible that expression aftereffects (and face aftereffects more generally) do not reflect adaptation of the mechanisms of perception, but instead reflect post-perceptual changes such as response biases (Storrs, 2015). If expression aftereffects are post-perceptual in origin, then adaptation studies may not be able to tell us anything about the way that expressions are processed at the perceptual level. Establishing that expression aftereffects have a perceptual locus is therefore crucial to our understanding of how expressions are visually represented.

In Chapter 4, we addressed this question by testing whether the timecourse of the expression aftereffect resembles the timecourse of perceptual aftereffects. There is a classic pattern of logarithmic build-up and exponential decay that has been demonstrated for lower-level aftereffects such as tilt (Wolfe, 1984), motion (Sekuler, 1975) and shape (Krauskopf, 1954). This timecourse pattern has also been demonstrated for identity and figural face aftereffects (Leopold, Rhodes, Müller, & Jeffery, 2005; Rhodes, Jeffery, Clifford, & Leopold, 2007). Importantly, it is difficult to explain how this timecourse pattern could be explained by post-perceptual processes such as response biases. We found that expression aftereffects follow this classic timecourse pattern. Our findings are
evidence that expression aftereffects are perceptual in nature. We therefore confirm that the visual representation of expression can adapt with experience, potentially helping to calibrate expression perception.

We extend this finding with a second experiment that demonstrates that relatively brief (seconds-long) periods of adaptation result in aftereffects that are surprisingly persistent given the dynamic nature of expressions. These persistent aftereffects may allow the visual system to integrate information over time, maintaining an updated norm that represents the statistics of recent experience.

Together, this work contributes to our understanding of the visual representation of facial expressions. We provide support for the opponent-coding of facial expression over an alternative, non-opponent model. We also find evidence that expression aftereffects are perceptual in nature, indicating that our expression coding mechanisms adapt based on recent experience. These aftereffects are persistent, which may allow the norm to integrate information about what is seen over time. Although we were unable to clearly distinguish between the two potential expression norms that have been proposed in the literature, we argue that either of these expressions may be useful as an approximation of the norm in most adaptation experiments.
References


McKone, E., Jeffery, L., Boeing, A., Clifford, C. W. G., & Rhodes, G. (2015). Corrigendum to "Face identity aftereffects increase monotonically with adaptor extremity over, but not beyond, the range of natural faces". *Vision research, 108*, 119-120.


Chapter Two

How is facial expression coded?
Abstract

Facial expression is theorized to be visually represented in a multi-dimensional expression space, relative to a norm. This norm-based coding is typically argued to be implemented by a two-pool opponent coding system. However, the evidence supporting the opponent coding of expression cannot rule out the presence of a third channel tuned to the center of each coded dimension. Here we used a paradigm not previously applied to facial expression to determine whether a central-channel model is necessary to explain expression coding. Participants identified expressions taken from a fear/anti-fear trajectory, first at baseline, and then in two adaptation conditions. In one condition, participants adapted to the expression at the center of the trajectory. In the other condition, participants adapted to alternating images from the two ends of the trajectory. The range of expressions that participants perceived as lying at the center of the trajectory narrowed in both conditions, a pattern which is not predicted by the central-channel model, but can be explained by the opponent-coding model. Adaptation to the center of the trajectory also increased identification of both fear and anti-fear, which may indicate a functional benefit for adaptive coding of facial expression.
How is facial expression coded?

We are able to extract many types of information from a face, including the person’s identity, gender, race, expression, and numerous other attributes. This ability requires sensitivity to small differences between faces, impressive given the similarity of faces as visual patterns. For instance, we are able to use subtle changes in the arrangement of the features of a face to perceive a person’s expression, which is often a source of important social cues. This sensitivity has led to great interest in the visual coding mechanisms that underlie the representation of facial expression and other types of facial information.

Many aspects of face perception (e.g. expression, identity, race, gender) are theorized to be visually coded relative to a norm in face-space. Each dimension in face-space represents a way in which faces are perceived to vary (although the dimensions used are not yet known). The norm represents the central tendency of previously-seen faces, and so lies at the center of this space. The positions of new faces in the space are coded relative to this norm. In this way, the coding of faces captures what is different or distinctive about them, which may contribute to our excellent face perception ability. Norm-based coding is also adaptive; the norm is constantly updated to represent the range of faces we encounter. This adaptability may help to calibrate the system to the range of faces that are most prevalent, optimizing sensitivity across that range (for a review see Rhodes & Leopold, 2011; Webster & MacLeod, 2011).

Evidence in support of norm-based coding of facial expression comes from paradigms using adaptation aftereffects (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Cook, Matei, & Johnston, 2011; Skinner & Benton, 2010, 2012). An aftereffect occurs when viewing a stimulus alters participants’ perception of subsequent stimuli. The responses of neural populations that were initially
stimulated by the stimulus begin to be suppressed with exposure. This reduction in responsiveness relative to other neural pools causes the percepts of subsequently viewed stimuli to be biased away from the adaptor. The greater the initial response, the larger the subsequent suppression and so the greater the aftereffect (Maddess, McCourt, Blakeslee, & Cunningham, 1988; Movshon & Lennie, 1979). Examining the size and direction of the aftereffects produced by adapting to particular expressions allows us to test hypotheses about the neural populations that code expression.

Norm-based face coding is often theorized to be instantiated by an opponent coding system (e.g. Rhodes & Jeffery, 2006; Rhodes et al., 2005; Robbins, McKone, & Edwards, 2007; Tsao & Freiwald, 2006). In this type of system there are two pools of neurons that code a given perceptual dimension, one pool that responds maximally to one extreme of the dimension and one pool that responds maximally to the other extreme. The norm is implicitly coded as the point at which the two pools respond equally. This model allows for efficient coding – the maximal response is reserved for aspects of a face that are particularly distinctive and so useful for recognition, with less energy devoted to representing the less useful aspects that are common to most faces. In the opponent coding model, adaptation is theorized to shift the position of the norm by altering the point where both neural pools respond equally. It is this shift in the norm that creates the aftereffect.

Evidence for the opponent coding of facial expression comes chiefly from the near-far aftereffect paradigm (Burton et al., 2013; Skinner & Benton, 2010, 2012). In the opponent coding model, adaptors that lie further from the norm will produce larger aftereffects than adaptors that lie closer to the norm (Robbins et al.,
This occurs because the more extreme adaptors result in stronger neural suppression and a larger shift in the norm than less extreme adaptors.

The adaptors used in the near-far paradigm are anti-expressions, produced by morphing a face along a trajectory that runs from an expression, through the average expression (a morph-average of the basic expressions, taken as an approximation of the norm, which is the central tendency of expressions that the participant has previously seen), and beyond the average to a point of equal distance beyond it. This anti-expression differs from the average to the same extent as the original expression, but in the opposite direction (so raised eyebrows become lowered, for instance). This expression/anti-expression trajectory does not necessarily correspond to an underlying coding dimension in face space. However, perception of faces along this trajectory will activate underlying expression-relevant dimensions, so we can use adaptation on the trajectory to examine the coding of those dimensions. Adapting to an anti-expression biases perception towards the original expression; for instance, adapting to anti-fear produces an aftereffect that biases perception towards fear (Burton et al., 2013; Skinner & Benton, 2010, 2012). As predicted by the opponent coding model, more extreme anti-expression adaptors create stronger aftereffects than less extreme anti-expression adaptors (Burton et al., 2013; Skinner & Benton, 2010, 2012).

The near-far paradigm may not, however, be sufficient to rule out alternative models with more than two channels. Narrowband multichannel models can produce an initial increase in aftereffects followed by a subsequent decline (Blakemore & Sutton, 1969; Clifford, Wenderoth, & Spehar, 2000). A central-channel model, with widely tuned channels coding the ends of the dimension but with an additional channel coding the center of the dimension (as
described by Calder, Jenkins, Cassel, & Clifford, 2008; Lawson, Clifford, & Calder, 2009, 2011), would also predict an initial increase in aftereffects. Therefore, we cannot currently rule out alternative, non-opponent models of expression coding.

There is evidence that a non-opponent multichannel model with a central channel, rather than opponent coding, is used to code gaze direction (Calder et al., 2008), and head and body orientation (Lawson et al., 2009, 2011). Those studies used a different paradigm, in which participants classified stimuli taken from along a trajectory using three labels: one for each end of the trajectory, and one for the center of the trajectory (e.g. “leftward gaze”, “rightward gaze” and “direct gaze” in the case of a gaze trajectory). Baseline judgments (i.e. without adaptation) are compared to judgments made in two adaptation conditions: one in which participants adapt to the center of the trajectory, and one in which participants adapt to alternating images of the ends of the trajectory. The dependent variable of interest is the range of stimuli that are judged to belong to the central category: the ‘central range’. If we assume that adaptation results in a suppression of neural pools stimulated by the adaptor, then we can derive predictions about the effect of adaptation on the central range from an opponent-coding model that differ from the predictions derived from a central-channel model.

In a central-channel model, alternating adaptation should narrow the central range, because the outer pools become relatively less responsive (Figure 1C). In contrast, central adaptation should widen the central range because the central channel becomes relatively less responsive (Figure 1E). This pattern was found for gaze direction and for head and body orientation. For instance, adapting to a central (front-facing) head direction reduced the range of directions perceived
as front-facing (narrowed the central range), while adapting to alternating left and right head directions increased the range of directions perceived as front-facing (widened the central range) (Lawson et al., 2011).

The opponent-coding model does not predict these opposing changes in the size of the central range. It is difficult to predict exactly what will happen to the size of the central range following adaptation in the case of opponent coding. Whether adaptation results in a narrowing or widening of the central range depends on the criterion by which a stimulus is perceived as central, and the shape of the response curves of the two pools (see Lawson et al., 2011 for further discussion). However, any change in the size of the central range should be in the same direction for both adaptation conditions. Alternating adaptation stimulates the two opponent pools, and central adaptation also stimulates these two opponent pools, although possibly to a lesser extent. For this reason, we can expect that any effect of adaptation seen in the alternating condition (whether widening or narrowing the central range) should also be seen in the central condition (Figures 1D and 1F). Thus in an opponent-coding model we do not expect the opposing changes in the size of the central range predicted by the central-channel model.
Figure 1. In a central-channel model (A), adapting to both endpoints of the dimension will increase the range of faces seen as central, shown here in grey (C); adapting to the center of the dimension will reduce the range of faces seen as central (E). In an opponent-coded model (B), adapting to both endpoints (D) will shift the central range in the same direction as adapting to the center (F). Figure adapted from Lawson et al. (2011).

We used this paradigm to determine which of the two models better describes the coding of facial expressions. We showed participants expressions from a morphed expression trajectory which ran from a fear expression, through the average expression, and out to an anti-fear expression. We chose this trajectory because fear and anti-fear are distinctive and would be easy for
participants to learn. We taught participants to use arbitrary labels to identify the expressions at each end of the trajectory (‘A’ for anti-fear and ‘C’ for fear) and the center of the trajectory (‘B’). Participants used these labels to classify faces taken from along the expression trajectory at baseline (no adaptation) and in two adaptation conditions: central adaptation, in which participants adapted to ‘B’, and alternating adaptation, in which participants adapted to alternating images of ‘A’ and ‘C’. Our dependent variable was the central range, the range of levels labeled ‘B’. Widening of the central range after alternating adaptation and narrowing of the central range after central adaptation (relative to baseline) would support a central-channel model. This opposing change cannot be explained by an opponent coding model, which predicts that any changes in the central range would occur in the same direction for both adaptation conditions.

Our experimental paradigm also allowed us to address an additional question – whether there is a functional benefit of adaptation for expression perception. In low-level vision, adaptive coding helps to calibrate the limited resources of the coding system to the current range of stimuli (Clifford, 2002; Clifford & Rhodes, 2005; Thompson & Burr, 2009), but evidence of this benefit in face perception has been mixed (for reviews, see Armann, Jeffery, Calder, Bülthoff, & Rhodes, 2011; Rhodes & Leopold, 2011). So far no research has examined a possible functional benefit of adaptive coding of facial expression.

Both the central-channel and opponent-coding models can accommodate an improvement in participants’ identification of expressions following adaptation. In the central-channel model, increased identification of fear and anti-fear following central adaptation can be explained by the suppression of the central neural channel (Figure 1E). In the opponent-coding model, identification of fear and anti-fear may be increased if adaptation steepens the tuning functions
of the two pools, altering their relative responses. Thus regardless of which model is supported by our results, we may find evidence of a functional benefit of expression adaptation.

**Method**

**Participants**

Twenty-four Caucasian participants were recruited from the University of Western Australia. This sample size was judged to be sufficient based on the size of samples used in research utilizing this paradigm with other stimuli (Calder et al., 2008; Lawson et al., 2009, 2011) and in research which has found significant aftereffects using expression stimuli (Burton et al., 2013; Skinner & Benton, 2010, 2012). One participant’s data were excluded from analysis due to a computer error during testing. The remaining 23 participants (5 male) had a mean age of 20.7 years, SD = 5.4 years. Participants were either awarded credit as part of a psychology course or were reimbursed $15 for travel expenses.

**Stimuli**

Stimuli were adapted from those used by Skinner and Benton (2010, 2012). These were gender-neutral expressive faces created from images of 20 Caucasian individuals (10 male and 10 female) posing various expressions. For each expression an average was created from the 20 images of that expression using morphing software (see Skinner and Benton, 2010 for more details). An overall average expression was created by taking an average of seven of these gender-neutral expressions (happy, sad, angry, fearful, surprised, disgusted and neutral).

The gender-neutral fear expression lies at one end of the test trajectory (‘100%’); the other end of the trajectory was created by morphing through the
average expression (‘0%’) and out into anti-fear (‘-100%’), which differs from the average to the same extent as fear but in the opposite direction (see Figure 2). The 100%, 0% and -100% expressions were our target expressions, also used as adaptors. The three expressions were labeled A (-100%), B (0%) and C (100%) for ease of participant response. The test stimuli were faces from nine points along the trajectory: -80%, -60%, -40%, -20%, 0%, 20%, 40%, 60% and 80% (Fig 3). To reduce testing duration the -100% and 100% expressions were not included as test faces.

Figure 2. The expressions defining the test trajectory, from left to right: -100% (antifear), 0% (average) and 100% (fear). For ease of participant response, these expressions were labeled A, B and C respectively.

Stimuli were shown in greyscale on a 21.5” iMac monitor, at a viewing distance of approximately 50cm. Adaptors subtended a visual angle of 8.9° x 12.1°. Test stimuli were shown at 75% of that size (6.9° x 9.1°) to reduce the contribution of retinotopic adaptation.
Figure 3. The nine test expression levels, ranging from anti-fear (-80%) through the average expression (0%) to fear (80%).

Procedure

Participants began each testing session with a training task that taught them to identify the three target expressions by their labels (A, B, C). Participants were first introduced to the expressions and their associated labels on screen. They were then shown an expression on screen and were given unlimited time to identify it using a marked keyboard key. A brief tone indicated whether the response was correct or not, and the next expression was shown. When participants were able to correctly identify a random sequence of nine expressions (each target expression appearing 3 times) they moved to the next training phase.

In the next training phase, participants saw the expressions for only 200ms. Each expression was followed by a 150ms blank ISI. Participants then saw a response screen (“?”), and responded as before. In this phase there was no feedback. Participants were again required to correctly identify a sequence of nine expressions to move on. If they went through three incorrect repetitions of this sequence they returned to the previous training phase and worked through that
again before coming back to the no-feedback training phase. Nine participants were required to return to the feedback phase in the first session, and eight participants were required to return to the feedback phase in the second session. Participants completed a mean of 2.7 total repetitions of the no-feedback training phase in the first session (SD = 2.1) and a mean of 2.3 total repetitions of the no-feedback training phase in the second session (SD = 1.7)$^1$.

The main testing procedure was adapted from Lawson et al. (2009). There were five testing phases – baseline, first adaptation, baseline, second adaptation, baseline. In the baseline phase participants were shown a test face for 200ms, followed by a 150ms ISI. Participants then identified which target expression they had seen using the labeled keys. Participants completed six of these trials for each of the nine test levels, in random order (54 trials total). In the alternating adaptation phase, participants first adapted to 40 alternating images of the -100% and 100% expressions (20 of each), each shown for 4000ms and separated by 200ms blank ISIs (total adaptation time of 160s). They then completed 54 trials of identifying test faces as in the baseline, but with each test face preceded by six alternating top-up adaptor images, each shown for 1000ms and separated by 200ms ISIs. The central adaptation condition followed the same procedure, but instead of the initial alternating adaptors participants adapted to the 0% expression for 40 repeated 4000ms exposures separated by 200ms blank ISIs (total adapt time of 160s), and the top-up images were six 1000ms exposures of the 0% expression separated by 200ms ISIs. Participants were allowed to move their eyes freely throughout the task.

$^1$ In the first session a computer error caused one participant to repeat the no-feedback training phase 46 times before finishing training. This participant has been left out of the training descriptives for the first session. In the second session they only required 5 repetitions to finish the training.

$^2$ This participant performed normally for the anger, disgust and fear expressions
To maintain attention during the long initial adaptation phase we included a secondary attention task. Over the course of some of the 4000ms-long adaptor exposures either the irises or lips would become brighter (see Fig. 4). This brightness change occurred in steps over the last 1000ms of the exposure: Five images were shown with the feature increasingly brightened; the first four brightness levels were shown for 125ms each, with the final, brightest image left on-screen for the remaining 500ms. Before adaptation began, participants were shown examples of the brightened eyes and lips, and were then shown a sequence of faces (4000ms exposures with 150ms ISIs) in which either the eyes or the lips changed as described above. Participants pressed a marked key as soon as they saw a change. Participants practiced this task until they and the experimenter were comfortable that they understood what was required. Of the 40 adaptation exposures eight contained an eye change and eight contained a lip change. Participants indicated whether the eyes or the lips changed using marked keys as soon as they saw a change. In the alternating adaptation phase changes were equally distributed across the -100% and 100% adaptors.
Figure 4. The brightness changes used in the attention task. From left to right: The original 0% expression; with the eyes brightened; with the lips brightened.

Participants began with two blocks of baseline testing (54 trials in each). They then learned to identify the eye and lip changes for the attention task. Next participants completed one block of adaptation testing (either alternating or central). This was followed by two more blocks of baseline testing. Participants were re-introduced to the eye and lip changes, and completed a second block of adaptation testing (whichever form they had not previously completed). Finally there were two more blocks of baseline testing.

Participants completed two of these testing sessions, each beginning with the training task. Participants either completed the alternating adaptation phase or the central adaptation phase first in both of their sessions; the order was counterbalanced between participants. Each session took approximately 45 minutes to complete. The two sessions were completed between 1 and 28 days apart (M = 6.48 days, SD = 5.35 days).

Results

Baseline data were collected in pairs of blocks before, between and after the two adaptation blocks. Following Lawson et al. (2009), we discarded the data from the first baseline block of each pair; from the first pair to allow a practice period, and from subsequent pairs to allow any lingering adaptation to dissipate as much as possible. The responses in the remaining baseline blocks (two, four and six) were taken as a measure of performance in the absence of adaptation. The two sessions of testing were collapsed together for analysis.

We aimed to determine whether a central-channel model is necessary for the coding of facial expression. To do this, we compared the range of expressions
which participants classified as central (‘B’) at baseline to the range classified as central in each of our two adaptation conditions. The central-channel model predicts that this range should widen relative to baseline after alternating adaptation, and narrow relative to baseline after central adaptation. The opponent coding model predicts that any changes in the range should be in the same direction for both adaptation conditions.

Mean proportions of A, B and C responses at each test strength in each adaptation condition are given in Figure 5. Examples of individual data from two participants are given in Figure 6. Visual inspection of the graphs indicates that the range of expression over which participants were more likely to respond “B” than “A” or “C” became narrower after adaptation in both conditions relative to baseline.

In order to statistically examine the effect of adaptation on the size of the central range we fit a mixed multinomial logit model to participants’ responses (see Supplementary Materials for details). This model uses the adaptation condition (baseline, alternating, central) to estimate the probability of selecting a given label (‘A’/‘B’/‘C’) in response to a given test level. The mixed logit model is an extension of standard logistic regression, but allows the parameters of the fit functions to vary across participants. This feature of the analysis makes it suitable for data like ours, in which multiple observations are taken from each participant. The resulting functions are presented in Figure 7.
Figure 5. Mean proportion of ‘A’ (anti-fear, shown in blue), ‘B’ (average, shown in red) and ‘C’ (fear, shown in green) responses to each level of the test trajectory in the baseline, alternating adaptation and central adaptation conditions, across all participants. To aid comparison across adaptation conditions, we show vertical lines that indicate the test level of the A-B and B-C crossing points in the baseline condition.
Figure 6. Examples of individual participants’ data: Mean proportion of ‘A’ (anti-fear, shown in blue), ‘B’ (average, shown in red) and ‘C’ (fear, shown in green) responses to each level of the test trajectory in the baseline, alternating adaptation and central adaptation conditions for two participants (left and right columns). To aid comparison across adaptation conditions, we show vertical lines that indicate the test level of the A-B and B-C crossing points in the baseline condition.
Figure 7. Response curves estimated by the mixed logit model. Curves show the probability of ‘A’ (anti-fear, shown in blue), ‘B’ (average, shown in red) and ‘C’ (fear, shown in green) responses to each level of the test trajectory in the baseline, alternating adaptation and central adaptation conditions. To aid comparison across adaptation conditions, we show vertical lines that indicate the test level of the A-B and B-C crossing points in the baseline condition. Points show mean proportion of responses from the group data, as plotted in Figure 5.
The points at which these functions crossed were determined for each adaptation condition. These crossing points represent the test level at which participants are equally likely to label an expression ‘B’ or ‘A’ and the test level at which participants are equally likely to label an expression ‘B’ or ‘C’. We compared the distance between these crossing points across the different adaptation conditions. The distance between the crossing points indicates the perceived central range – the range of test levels that participants tended to label as ‘B’ more than ‘A’ or ‘C’. As can be seen in Figure 7, the central range of our fit functions narrowed after alternating adaptation compared to baseline, and also narrowed after central adaptation compared to baseline. We estimated the standard error of these shifts, allowing us to test their significance, using the nlcom function of the STATA data analysis package (StataCorp, 2013), which uses the delta method (Oehlert, 1992). The inter-threshold distance was significantly narrower than it was at baseline in both the alternating and central conditions (Table 1). Importantly, the range changed in the same direction in both conditions, supporting the opponent coding model.
Table 1. The range of test levels judged as central (‘B’) at baseline and in each adaptation condition, and tests of the significance of the changes in this range from the baseline condition. Range is shown in units of adaptor strength percentage.

<table>
<thead>
<tr>
<th>Adaptation condition</th>
<th>Central range</th>
<th>Change in central range from baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>Baseline</td>
<td>82.74</td>
<td>2.26</td>
</tr>
<tr>
<td>Alternating</td>
<td>76.55</td>
<td>2.88</td>
</tr>
<tr>
<td>Central</td>
<td>62.83</td>
<td>1.94</td>
</tr>
</tbody>
</table>

^a Cohen's d was calculated here by dividing the difference between the baseline and post-adapt values by the pooled standard deviations of the baseline and post-adapt values (Dunlap, Cortina, Vaslow, & Burke, 1996). Standard deviations were produced by multiplying the approximated standard errors by the square root of N.

Our second aim was to investigate the possible functional benefit of expression adaptation. The narrowing of the central range observed above indicates that participants were more likely to identify fear and anti-fear around the average in both adaptation conditions. However, there was a greater narrowing of the central range for central adaptation compared to alternating adaptation, z = -4.96, p < .001, d = 0.74, indicating that central adaptation increased identification of subtle expressions more than alternating adaptation did. To further examine the effect of adaptation on participants’ identification of fear and anti-fear we compared their responses to the strongest test expressions, -80% anti-fear and 80% fear, between adaptation and baseline. As can be seen in Fig. 7, central adaptation made the -80% expression more likely to be labeled ‘A’ (anti-fear), z = 9.60, p < .001, d = 1.64, and the 80% expression more likely to be labeled ‘C’ (fear), z = 10.60, p < .001, d = 1.82. Alternating adaptation did not significantly
change the likelihood of identifying the -80% or 80% expressions as ‘A’ and ‘C’ respectively, $z = -1.07, p = .283, d = 0.21$ and $z = 1.71, p = .088, d = 0.28$ respectively. Therefore, although both adaptation conditions increased identification of subtle expressions around the average, only adaptation to the average expression increased identification of the stronger expressions at the ends of the trajectory. Again, this suggests that adapting to the average had a greater effect on identification of expressions than adapting to the extremes of the trajectory.

Alternating adaptation also caused the functions describing participant responses to shift towards the ‘fear’ end of the trajectory, indicating a bias to see faces as less fearful in this condition. Both the A-B threshold and B-C thresholds shifted significantly towards ‘fear’ compared to baseline, $z = 4.90, p < .001, d = 1.07$, and $z = 2.02, p = .043, d = 0.44$ respectively. In the central condition, both the A-B and B-C thresholds shifted significantly inward, towards 0%, $z = 8.82, p < .001, d = 1.76$, and $z = 4.94, p < .001, d = 1.13$, respectively, indicating that there was no overall bias away from ‘fear’ responses in this condition. These results indicate that there may be an imbalance in the size of the aftereffects produced by fear and anti-fear.

In the analyses above we compared the effects of adaptation between the central and alternating conditions. It is important to be sure that participants attended equally to the adaptors in both conditions, as better-attended adaptors tend to produce larger aftereffects, (Rhodes et al., 2011). To check that participants were not attending more to the adaptor in the central condition than in the alternating condition, we looked at performance in the change detection task that took place during the long period of initial adaptation. We calculated the proportion of eye changes and lip changes that were correctly identified for each
of the three adaptors (0%, central condition, and -100% and 100%, alternating condition) (Table 2). A repeated-measures ANOVA revealed no significant main effect of adaptor, $F(2,44) = 0.11, p = .896, \eta^2_p = .01$, a significant main effect of feature, $F(1, 22) = 4.28, p = .050, \eta^2_p = .16$, and no significant interaction, $F(2,44) = 1.59, p = .215, \eta^2_p = .07$. Thus although participants were significantly better at detecting eye changes than lip changes, there was no difference in their performance across the three adaptors. This finding indicates that the differences we found between the central and alternating adaptation conditions were not due to differences in attention.

Table 2. The proportion of eye and lip changes correctly identified for each adaptor (-100% anti-fear, 0% average and 100% fear) and overall proportion of changes detected for each adaptor during the change detection task.

<table>
<thead>
<tr>
<th>Adaptor</th>
<th>Eye changes</th>
<th>Lip Changes</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SE$</td>
<td>$M$</td>
</tr>
<tr>
<td>0%</td>
<td>.76</td>
<td>.04</td>
<td>.63</td>
</tr>
<tr>
<td>-100%</td>
<td>.72</td>
<td>.04</td>
<td>.71</td>
</tr>
<tr>
<td>100%</td>
<td>.75</td>
<td>.04</td>
<td>.66</td>
</tr>
</tbody>
</table>

**Discussion**

We investigated whether a central-channel model, rather than a pure opponent model, is necessary to explain the coding of facial expression. To do so we compared the effects of adapting to either alternating extreme adaptors (fear and anti-fear) or a central adaptor (average) on the range of expressions seen as central on a fear-anti-fear expression trajectory. We found a significant inward shift in the central range for both the alternating and central adaptation conditions.
relative to baseline. This pattern of results is not predicted by a central-channel model. It can, however, be accommodated by an opponent-coded model.

Our findings agree with several other studies that have found evidence for the opponent coding of facial expression (Burton et al., 2013; Skinner & Benton, 2009, 2010). These previous studies have used the near-far paradigm, which compares the sizes of the aftereffects produced by weak and strong anti-expressions. Stronger anti-expressions produce larger aftereffects, indicating norm-based coding. However, this paradigm does not necessarily distinguish between an opponent coding model and certain multi-channel models. The present study used a different paradigm not previously applied to facial expression, and provides converging evidence in support of the opponent coding of expression.

As well as perceptual adaptation, changes in decision-making may have contributed to the shifts in responses that we see here. By the nature of this paradigm, the adaptors are also the exemplars of the three expression categories. Seeing these expressions during the adaptation blocks may have provided a reference that helped to reduce uncertainty in participants’ choices, or changed the criteria by which they made those choices. In future this potential referencing effect could be reduced (although not eliminated) by using different identities for adapt and test. However, there is evidence that a portion of the expression aftereffect is identity-dependent (Skinner & Benton, 2012), so future research taking this measure would need to take into account an expected decrease in the size of the aftereffects.

Our results may indicate a functional role of adaptation. The narrowing of the central range in both adaptation conditions relative to baseline indicates an increase in identification of fear and anti-fear after adapting to faces from the trajectory. Adaptation may help to calibrate perceptual resources to recently-seen
stimuli, improving discrimination where it is needed. This functional benefit of adaptation has previously been supported in the perception of other facial attributes such as identity (Rhodes, Watson, Jeffery, & Clifford, 2010), gender (Yang, Shen, Chen, & Fang, 2011), and viewpoint (Chen, Yang, Wang, & Fang, 2010), but has not previously been established for facial expression.

We should be cautious, however, about interpreting increases in expression identification following adaptation as an improvement in sensitivity. Again, these effects could also be explained by changes in participants’ decision criteria in the adaptation conditions. Viewing the adaptors may give participants a reference point against which to judge the test faces, altering their responses. For instance, if a participant tended to respond ‘B’ when uncertain, reducing that uncertainty by providing a reference would improve their identification rates for fear and anti-fear expressions. Any potential functional benefit of expression adaptation should be investigated in future research using a measure of sensitivity that is less affected by response biases, such as an odd-man out task (O'Mahony, 1995).

It is interesting to consider that we were able to produce aftereffects in our participants after adapting them to the average expression, which is an approximation of the norm. In norm-based models, aftereffects are typically explained as the result of shifts in the position of the norm produced by adaptation. Adaptation to the norm itself should not produce such a shift, and so might not be predicted to produce an aftereffect. However, in the opponent-coded instantiation of norm-based coding supported by our findings, changing the position of the norm may not be the only way to produce an aftereffect. Adaptation may also change the slope of the tuning functions of the opponent pools, so that the relative response rate of the two pools to a given stimulus may
be altered even when the norm, the point at which they respond equally, remains the same. This kind of change in response can be visualized as a distortion of expression space that stretches the space around the norm, making expressions around that point appear more distinctive.

It should be noted that, following Calder et al. (2008) and Lawson et al. (2009, 2011), our initial predictions are based on the assumption that adaptation suppresses the neural channels stimulated by the adaptor. If adaptation also affects the shape of the response curves, it would make predictions about the result of adaptation in each model more complex. In the case of the opponent-coding model, the simplest situation is one where adaptation in both the central and alternating conditions affects the shape of the response curves in the same way. If this is the case, our initial prediction (that both alternating and central adaptation will have the same effect on the size of the central range) remains the same. It is also possible (but less plausible) that adaptation in one condition might steepen the response curves, while adaptation in the other condition might flatten them. If this were the case, the opponent-coding model might be able to accommodate opposing changes in the size of the central range. The central-channel model is even more complex, as there are three or more neural channels to consider, and the response curve of the central channel may be affected by adaptation differently to that of the outer channels. Computer simulation of the models may be helpful for determining the predicted effects of this adaptation paradigm as we vary our assumptions (cf. Ross, Deroche, & Palmeri, 2013 for an example of how this might be approached).

We chose an average expression as our approximation of the norm, as has been done in several other expression aftereffect studies (Burton et al., 2013; Cook et al., 2011; Skinner & Benton, 2010, 2012). We chose this expression
because the norm represents the central tendency of previously seen faces (Valentine, 1991). However, other studies have instead used a neutral expression as their norm (Juricevic & Webster, 2012; Rutherford, Chattha, & Krysko, 2008; Rutherford, Troubridge, & Walsh, 2011). The neutral expression represents the face at rest, which could be argued to make it a good, ‘unexpressive’ candidate for the norm, analogous to white in colour-space (Juricevic & Webster, 2012). However, there is evidence that the neutral face is not actually perceived as expressively neutral. In an implicit emotion evaluation task Lee, Kang, Park, Kim, and An (2008) found that participants responded to neutral stimuli in the same way they did negative stimuli. Additionally, in studies that map the perceptual organization of emotion based on participants’ judgments of expressions, neutral expressions are often located away from the center of the space, nearer to other prototypical emotion expressions (Bimler & Kirkland, 2001; Gao, Maurer, & Nishimura, 2010; J. A. Russell & Bullock, 1985; J.A. Russell & Bullock, 1986). An average expression may therefore be the more appropriate choice.

We found an overall bias towards ‘anti-fear’ responses in the alternating adaptation condition. This bias suggests an imbalance in the amount of adaptation produced by the 100% and -100% expressions, perhaps because the fear expression is more familiar or more attention-grabbing than anti-fear. Better-attended adaptors tend to produce larger aftereffects (Rhodes et al., 2011). However, there was no difference in change detection performance between the two adaptors, so any difference in attention was not large enough to affect performance this task. The shape and textural information of the -100% anti-fear expression differ from the average to the same extent as those of the 100% fear expression. Nevertheless, it is possible that 100% fear is more perceptually dissimilar to the average expression than -100% anti-fear, which would result in
stronger adaptation to the fear end of the trajectory (Robbins et al., 2007). This imbalance is not a problem for the interpretation of our data, however, because we still found a significant reduction of the range of expressions seen as central following alternating adaptation as well as following central adaptation.

Although we are using a fear-antifear trajectory here to draw conclusions about the coding systems that underlie expression perception, we do not assume that fear-antifear is an explicitly coded dimension of expression space, or that there are specific anti-fear detectors that are affected by adaptation to anti-fear. Rather, like previous studies (Burton et al., 2013; Skinner & Benton, 2010, 2012), we assume that adaptation along the fear-antifear trajectory adapts a number of underlying, expression-relevant dimensions. These dimensions might relate to the positions of particular features or muscle groups, or could describe key components of the statistical variation found in facial postures (Cook et al., 2011). By studying participants’ judgments of the composite fear-antifear trajectory we can indirectly observe the adaptation of those underlying dimensions.

Our method of data analysis differed from that of Lawson et al. (2009, 2011) and Calder et al. (2008). In those studies five test levels were used (e.g. 10° left, 5° left, direct, 5° right and 10° right for gaze direction) and changes to the central range were approximated by calculating changes in the proportion of expressions judged to be central at levels immediately to either side of the center of the trajectory (e.g. 5° left and 5° right in the above example). An increase in the proportion judged to be central at these levels indicated a widening of the central range; a decrease at these levels indicated a narrowing of the central range. As the alternating condition in our study caused an overall bias in responses towards ‘fear’, this simple analysis was not appropriate for our data. Fitting the mixed multinomial logit model allowed us to more directly test the predicted changes in
the central range, and to quantify the size of these changes. It should be noted that for mixed logit models, there is no intuitive measure of goodness of fit (Train, 2003). The log likelihood ratio indicates that our model fit better than the unconstrained model, but unlike an $R^2$ this measure cannot be interpreted in terms of the proportion of variance explained. However, the closeness of the fitted functions (Figure 7) to the pattern of mean proportions of responses (Figure 5) confirms that the model is an appropriate representation of the pattern of the data.

In the present study we used a single expression trajectory, running from fear to anti-fear. Testing one trajectory allowed us to maximize the power of our analysis. Expressions are modeled as being coded in a single, multi-dimensional expression-space, so the coding system that applies to the fear-anti-fear trajectory should be common to other expression trajectories. However, in future it may be useful to test other expression trajectories to ensure that our findings generalize across multiple expressions. In particular, use of a trajectory that runs between two unfamiliar expressions (such as the trajectories defined by primary components analysis in Cook et al., 2011) may prevent the asymmetric adaptation created by using a familiar expression and unfamiliar anti-expression as the endpoints of the test trajectory.

Future research should seek converging evidence that expression is opponent-coded by applying other paradigms that can distinguish between opponent and multi-channel models to expression stimuli. One such paradigm, used previously to examine the coding of gender in faces (Pond et al., 2013), involves adapting participants to a very wide range of adaptor strengths. This approach improves on the near-far paradigm by using adaptors which extend beyond the natural range of facial variation, where we are more likely to capture differences between opponent and multi-channel coding. If, as suggested here,
facial expression is opponent-coded, we would expect expression aftereffects to increase with increasing adaptor strength across the natural range of variation, and then either continue to increase or asymptote beyond that range.

Our present findings support an opponent coding model of facial expression, rather than a central-channel model. They are in line with previous studies that have found evidence supporting the opponent coding of facial expression using the near-far paradigm (Burton et al., 2013; Skinner & Benton, 2010, 2012). We also found that adaptation to a range of expressions, and particularly the central tendency of that range, resulted in increased identification of the fear and anti-fear expressions, possibly indicating a functional benefit of adaptation for expression perception.
Acknowledgements

This research was supported the UK Medical Research Council under project code MC-A060-5PQ50 (Andrew J. Calder), the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021), an ARC Professorial Fellowship to Rhodes (DP0877379), and an ARC Discovery Outstanding Researcher Award to Rhodes (DP130102300). We thank Andrew Skinner and Chris Benton for providing the stimuli, Michael Burton for assistance with data analysis, and Colin Clifford for helpful discussions. Ethical approval was granted by the Human Research Ethics Committee of the University of Western Australia. This paper is dedicated to the memory of Andy Calder (1965-2013).
References


O'Mahony, Michael. (1995). Who told you the triangle test was simple? *Food
Quality and Preference, 6*(4), 227-238.

46*(1), 27-29.

Aftereffects support opponent coding of face gender. *Journal of Vision,
13*(14).

*Vision research, 46*, 2977-2987.

Rhodes, G., Jeffery, L., Evangelista, E., Ewing, L., Peters, M, & Taylor, L.
1811-1819.

identity. In A. J. Calder, G. Rhodes, M. H. Johnson & J. V. Haxby (Eds.),
*The Oxford Handbook of Face Perception* (pp. 263-286). Oxford: Oxford
University Press.

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 Aftereffects in High-Level Vision


StataCorp. (2013). Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.


Supplementary Material: The Mixed Multinomial Logit Model

The mixed logit model is a discrete choice model widely used in areas such as transport, health and environmental economics. Like a simple logit model, the mixed logit model estimates the probability of a choice based on the assumption of an underlying latent variable that is maximized (generally ‘utility’ in economics, but potentially generalizable). With unlabeled alternatives the latent is defined as:

\[ y_{ni}^* = \beta' x_{ni} + \epsilon_{ni} \]

Where \( y^* \) is the latent index associated with alternative \( i \) by person \( n \), \( \beta' x_{ni} \) is a linear function that defines the relationship between the observed factors \( (x_{ni}) \) and the latent, and \( \epsilon_{ni} \) is a random component that is unobservable to the researcher.

The definition of the latent above gives a single utility function, and assumes that the observed factors vary between alternatives. In the case of labeled alternatives, parameter values may vary across alternatives while levels of the observed factors do not. This is the case here, where we are interested in estimating the probability of selecting each alternative subject to a common test level. We therefore require a definition of the latent that estimates different parameters for each alternative, while keeping the observed factors constant for all alternatives:

\[ y_{ni}^* = \beta_i' x_n + \epsilon_{ni} \]

The vector of parameters \( (\beta_i') \) that describes the relationship between the observed factors and the latent is unknown and must be estimated from the
observed responses. The standard logit model assumes that the parameters are the same across all individuals. If we assume that the random component is IID and follows a Gumbel distribution, and that the individual selects the alternative that yields the highest $y^*$, then we arrive at the logit formulation. The probability of a choice is given by:

$$P_{ni} = \frac{e^{\beta_i x_n}}{\sum_j e^{\beta_j x_n}}$$

Where participant $n$ makes choice $i$ from choice set $j$.

The mixed logit model relaxes the assumption that all individuals have the same parameters in the linear function $\beta_i x_n$, by allowing individual-specific estimates. It is assumed that the individual-specific parameters are drawn from the ‘mixing distribution’, density $f(\beta|\theta)$, and that they are constant for an individual across all of their choice situations. $\theta$ are the parameters that describe the density of $\beta$. The probability of a choice is estimated by taking a weighted average of the logit formula at each set of parameter values, where the weights are given by $f(\beta|\theta)$. The probability of participant $n$ making choice $i$ becomes:

$$P_{ni} = \frac{e^{\beta_{ni} x_n}}{\sum_j e^{\beta_{nj} x_n}} f(\beta|\theta) d\beta$$

In this way the mixed logit model can account for systematic differences between participants in panel data, as different parameters are allowed for each participant. In estimating the model, one identifies the parameters ($\theta$) describing the mixing distribution (typically the means, standard deviations and covariances of the
random distributions) from which the individual marginal utilities are drawn. All
or some of the parameters of the utility function may be specified as random.

In our data, participants made a choice from three alternatives (A, B and C). Leaving adaptation condition aside for the moment, the probability of making a choice can be predicted from the test level in that trial, $x$. We therefore have three linear functions describing the relationship between $x$ and the latent ($\beta_{ni}x_n$) :

$$\beta_{nA}x = b_{xA}x + a_{nA}$$
$$\beta_{nB}x = b_{xB}x + a_{nB}$$
$$\beta_{nC}x = b_{xC}x + a_{nC}$$

Where $\beta_{nA}x$ is the linear function for alternative A, and so on. For identification purposes the parameters of one alternative must be set to zero: we selected alternative B.

Individual participants made multiple choices, and it is possible that each participant had a consistent bias in their probability of choosing each alternative across all of their decisions (for instance, an overall bias to select B more often than the average participant). The mixed logit model allows us to deal with this by making the intercept terms $a_{nA}$ and $a_{nC}$ random parameters that can vary between participants; this is equivalent to having a fixed constant and an additional random error with mean zero that captures consistent differences between participants’ evaluation of alternatives. We assume that the random effects are normally distributed, and potentially correlated with one another.

We then expand the model to include the three adaptation conditions. Condition is modeled as having an impact on both the slope coefficient of $x$ and on the mean of the intercepts, allowing condition to affect the probability of
selecting each alternative. For ease of notation, we will term the alternating adaptation condition t1, the baseline condition as t2 and the central adaptation condition as t3. We used the parameters described above for the alternating adaptation (t1) condition:

\[ \beta_{nA}^{t1}x = b_A x + a_{nA} \]
\[ \beta_{nC}^{t1}x = b_C x + a_{nC} \]

For the baseline (t2) and central adaptation (t3) conditions, two additional fixed parameters were added for each alternative to adjust the slope and mean intercept:

\[ \beta_{nA}^{t2}x = (b_A + bt2_A)x + a_{nA} + at2_A \]
\[ \beta_{nC}^{t2}x = (b_C + bt2_C)x + a_{nC} + at2_C \]

\[ \beta_{nA}^{t3}x = (b_A + bt3_A)x + a_{nA} + at3_A \]
\[ \beta_{nC}^{t3}x = (b_C + bt3_C)x + a_{nC} + at3_C \]

The random effect for each participant remains constant across all three conditions.

The analysis was run using the mixlogit function of the Stata statistics package (StataCorp, 2013). The estimated parameters are given in Table S1. The functions described by these parameters are graphed in Figure 7. These functions were used to calculate our dependent variables as described in the Results section.
For a mixed logit model, goodness of fit can be measured by the likelihood ratio index, which compares the simulated log-likelihood of the complete model to the simulated log-likelihood of the model with all parameters set to zero (McFadden, 1974). The index can take values between 0 (chance performance) and 1 (perfect performance). For the present model, $R^2_M = 0.45$.

Our model combines data from three separate baseline periods into a single condition. This assumes that baseline responses remain the same across the testing periods. To check this assumption, we compared the fit of two mixed logit models to the baseline data (disregarding the data from the adaptation conditions). The simplest model treated all of the baseline data as belonging to single condition. The alternative model divided the baseline data into three conditions: an initial baseline condition, a post-alternating-adaptation baseline condition, and a post-central-adaptation baseline condition (Figure S1). If responses differ between the three baseline conditions, then the second model should fit the data better than the first. To test the relative fit of the models we used a likelihood ratio test, which compares the log likelihoods. The test statistic is distributed as a chi-square, with degrees of freedom equal to difference in the degrees of freedom between the two models.

For the single condition model, log likelihood = -4791.148, df = 4. For the three conditions model, log likelihood = -4789.822, df = 16. This gives $D(12) = 2.652$, $p = .998$. This indicates that allowing separate curves to be fit for the three baseline periods does not improve the fit compared to a model that treats them as a single condition. We can conclude that the responses do not differ significantly between the three baseline periods.
Table S1. Parameters estimated by the mixed logit analysis. For random
parameters we estimate the mean and variance of the random distribution. The
variance of $a_{nA}$ is given here as $\sigma^2_A$, and so on.

|                  | Coef.  | Std. Err | z     | P>|z|  | 95% Conf. Interval |
|------------------|--------|----------|-------|-------|-------------------|
| **Fixed Parameters** |        |          |       |       |                   |
| $b_A$            | -0.034 | 0.002    | -20.44| 0.000 | -0.037            | -0.031            |
| $b_C$            | 0.055  | 0.003    | 21.98 | 0.000 | 0.050             | 0.060             |
| $bt_{2A}$        | -0.012 | 0.002    | -5.74 | 0.000 | -0.016            | -0.008            |
| $bt_{3A}$        | -0.026 | 0.003    | -8.42 | 0.000 | -0.032            | -0.020            |
| $bt_{2C}$        | -0.011 | 0.003    | -3.99 | 0.000 | -0.016            | -0.006            |
| $bt_{3C}$        | 0.025  | 0.004    | 5.77  | 0.000 | 0.017             | 0.034             |
| $at_{2A}$        | -0.809 | 0.097    | -8.33 | 0.000 | -1.000            | -0.619            |
| $at_{3A}$        | -0.634 | 0.131    | -4.85 | 0.000 | -0.891            | -0.378            |
| $at_{2C}$        | 0.622  | 0.127    | 4.91  | 0.000 | 0.374             | 0.870             |
| $at_{3C}$        | -0.281 | 0.180    | -1.56 | 0.119 | -0.635            | 0.072             |
| **Random Parameters** |        |          |       |       |                   |
| Mean $a_{nA}$    | -1.102 | 0.087    | -12.73| 0.000 | -1.271            | -0.932            |
| Mean $a_{nC}$    | -2.446 | 0.138    | -17.74| 0.000 | -2.716            | -2.176            |
| **Variance Covariance Estimates for Random Parameters** | | | | | |
| $\sigma^2_A$    | 1.018  | 0.113    | 8.98  | 0.000 | 0.796             | 1.241             |
| $\sigma^2_C$    | 0.979  | 0.146    | 6.70  | 0.000 | 0.692             | 1.265             |
| $cov_{AC}$      | -0.779 | 0.077    | -10.06| 0.000 | -0.931            | -0.627            |
Figure S1. Functions fit by the multinomial logit model for the three baseline periods (Baseline 2, before adaptation; Baseline 4, after alternating adaptation; Baseline 6, after central adaptation), showing probability of responding A, B and C to each test level. Points indicate mean proportion of A, B and C responses at each test level.
References


StataCorp. (2013). Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.
Chapter Three
Is a Neutral or an Average Expression Closer to the Norm of Expression Space?
Abstract

There is evidence that facial expressions are visually represented in a multi-dimensional space, relative to a central norm expression. However, there is no consensus on which expression serves as this norm. We considered two contenders for this position: the neutral expression, and the average expression. We compared the size of expression aftereffects produced by adaptation on trajectories that passed through either a neutral or an average face. Previous research has shown that adaptation on trajectories that pass through the norm produces larger aftereffects than adaptation on trajectories that do not pass through the norm. We found no difference in aftereffect size between adaptation on the two sets of trajectories, offering no evidence for one contender over the other. Both the average and neutral expressions appear relatively unexpressive, and both may be reasonably good approximations of the norm. We found significant aftereffects for both sets of trajectories, indicating that either expression may be useful as an approximate norm in most expression adaptation research.
Is a Neutral or an Average Expression Closer to the Norm of Expression Space?

Facial expressions are an important source of social information, providing cues to the emotions and intentions of others. Despite the similarity of faces as visual patterns, most people are able to quickly and easily interpret expression from a single glance at a face. Norm-based coding is one mechanism that may contribute to our proficiency in making judgments about facial expressions (Rhodes & Leopold, 2011). In this type of visual coding, faces are represented in a multi-dimensional space in which each dimension represents a way in which faces are perceived to vary. The ‘norm’ is located at the center of the space, and acts as a reference point against which the positions of other faces are coded. This type of coding is efficient, because it focuses coding on what is diagnostic about a face rather than what is common to most faces, and flexible, because the system is calibrated by experience (Rhodes & Leopold, 2011; Webster & MacLeod, 2011).

Adaptation aftereffects are a useful tool for investigating how facial expressions are visually represented. These aftereffects occur when exposure to one stimulus reduces the responsiveness of the neurons that code it, altering the neural response elicited by subsequent stimuli (Webster & MacLeod, 2011). For instance, adapting to constant motion in one direction will temporarily make a stationary object viewed immediately afterwards appear to move in the other direction (Anstis, Verstraten, & Mather, 1998). Adaptation can be induced by viewing faces: for example, adapting to faces that have been distorted so that the features are contracted into the center of the face will make an undistorted face appear expanded (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Webster & MacLin, 1999), and adapting to male faces makes an androgynous
face appear more female (Webster, Kaping, Mizokami, & Duhamel, 2004). These face aftereffects are likely to reflect adaptation of higher-level face representations, not just low-level, retinotopic adaptation, because they persist over changes in the size, colour, contrast and orientation of a face, and when free eye movements are allowed (Rhodes et al., 2005; Watson & Clifford, 2003; Yamashita, Hardy, De Valois, & Webster, 2005; Zhao & Chubb, 2001).

Similarly, adaptation to a particular facial expression will bias perception of faces seen after adaptation towards the opposite expression (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Cook, Matei, & Johnston, 2011; Juricevic & Webster, 2012; Skinner & Benton, 2010, 2012). The opposite of an expression is its anti-expression, which has attributes that differ from the norm to the same extent as the original expression, but in the opposite direction. For instance, where fear has raised brows relative to the norm, anti-fear has lowered brows (Figure 1). These anti-expressions are often used as the adaptors in expression adaptation experiments. Adapting to an anti-expression makes faces seen after adaptation look more like the original expression (Burton et al., 2013; Cook et al., 2011; Juricevic & Webster, 2012; Skinner & Benton, 2010, 2012). These effects persist over changes in size and when free eye movements are allowed (Burton et al., 2013; Cook et al., 2011; Fox & Barton, 2007; Hsu & Young, 2004; Skinner & Benton, 2010, 2012), again suggesting that they tap higher-level face processing.
Evidence from several aftereffect studies suggests that facial expressions are coded relative to a norm (Burton et al., 2013; Cook et al., 2011; Skinner & Benton, 2010, 2012). Norm-based coding can be implemented by a two-pool opponent coding system. In this type of coding, two pools of neurons code each dimension of perceived expression variation, one responding maximally to each end of the dimension. The norm is represented by the point where both pools respond equally. Converging evidence from a range of paradigms supports a two-pool coding system that represents facial expressions relative to a norm (Burton, Jeffery, Calder, & Rhodes, 2015; Burton et al., 2013; Cook et al., 2011; Skinner & Benton, 2010, 2012).

Although this evidence suggests that a norm expression plays an important role in the representation of facial expressions, there is currently no consensus on precisely what this norm expression might be. Studies investigating norm-based coding of facial expressions have used one of two different potential expression norms: an average expression (the central tendency of facial expressions), or a neutral expression (when the face is at rest) (Figure 2). The average expression is

**Figure 1.** Fear (left) and anti-fear (right). The feature positions of these two expressions are opposite one another relative to the central tendency of facial expressions (center).
made by taking an average from a set of expression images, of either the basic emotion expressions (Burton et al., 2013; Skinner & Benton, 2010, 2012) or more varied facial postures (Cook et al., 2011). This average expression is intended to lie near the mean on each of the various dimensions on which expressions are coded. An average expression is analogous to the type of norm often used in facial identity research (where the concept of face space was originally developed), which is generally created by averaging together a large set of different identities (for a review see Rhodes & Leopold, 2011). For expression coding, the benefit of an average as the norm is that the average represents what is common to most expressions, maximizing the efficiency granted by norm-based coding.

Other researchers have used a neutral expression as the expression norm (Juricevic & Webster, 2012; Rutherford, Chattha, & Krysko, 2008; Rutherford, Troubridge, & Walsh, 2011). Webster and MacLeod (2011) suggest that the neutral expression has the property of appearing unexpressive or “psychologically neutral,” analogous to the way that white, the norm of opponent-coded colour space, appears neutral. Indeed, it is commonly assumed that a neutral face is unexpressive, and so can be used as a baseline against which other expressions are compared. For example, tests of expression sensitivity create expressions of varying strengths by morphing between a given expression and neutral (e.g. Gao & Maurer, 2009, 2010). Likewise, the neutral expression has been used to gauge the baseline response in neural imaging studies looking at responses to facial expressions (Sprengelmeyer et al., 1998; Kesler-West et al., 2001; Pessoa et al., 2002; Kilts et al., 2003).

Not all evidence, however, suggests that the neutral face is unexpressive. In an implicit emotion evaluation task, Lee, Kang, Park, Kim, and An (2008) found that participants responded to neutral expression stimuli in the same way
they did to negative expression stimuli. Moreover, multidimensional scaling of facial expressions in several studies has placed neutral expressions away from the center of the space, nearer to other emotion expressions, such as sadness or sleepiness (Bimler & Kirkland, 2001; Gao, Maurer, & Nishimura, 2010; J. A. Russell & Bullock, 1985; J. A. Russell & Bullock, 1986). These findings suggest that the “neutral” expression might be better regarded as another expression, with negative rather than neutral valence. If so then this undermines the principle reason to believe that it should be the norm of expression space.

Our aim was to determine more directly which expression, average or neutral, is a better approximation of the norm of expression space. To do so, we exploited the finding that in a norm-based system, adapt-test pairs that lie opposite one another relative to the norm produce larger aftereffects than adapt-test pairs that are not opposite one another (Leopold, O'Toole, Vetter, & Blanz, 2001), even when perceptual distance between the pairs is equal (Rhodes & Jeffery, 2006). This effect occurs because adaptation selectively biases perception towards the opposite of the adaptor across the norm. This property has been used to identify the best approximation of the norm for different face categories in facial identity research. For example, adapt-test pairs that lie opposite one another relative to a race-specific average (Asian or Caucasian) produce larger identity aftereffects than adapt-test pairs opposite one another relative to a mixed-race average, indicating that the norms used to code identity are race-specific (Armann, Jeffery, Calder, Bülthoff, & Rhodes, 2011). Similarly, adapt-test pairs that are opposite on another relative to a sex-specific average produce larger identity aftereffects than adapt-test pairs opposite one another relative to an androgynous average, indicating that identity is coded relative to sex-specific norms (Rhodes, Jaquet, et al., 2011). Following this logic, expression aftereffects should be larger on
trajectories that pass through the true expression norm (opposite trajectories) than on trajectories that do not (non-opposite trajectories). Therefore, if one of our two potential norms (neutral or average) is a better approximation of the true expression norm, then we will see larger aftereffects on trajectories that pass through that expression (with adapt-reference face pairs equated on perceptual similarity).

We compared adaptation on trajectories that pass either through an average expression, or through a neutral expression. We refer to these two potential approximations of the norm as ‘reference faces’ because we used them as reference points around which we created expression trajectories. Participants viewed a range of faces taken from the same trajectory as each adaptor, and labeled the expressions that they saw on those faces both before and adaptation. Adaptation should make a test face appear more like the expression opposite the adaptor. We measured the aftereffect as the increase in participants’ tendency to respond with this opposite expression following adaptation, relative to their pre-adaptation responses. We expected that if one of the two reference faces was a better approximation of the norm expression than the other, then we should see larger aftereffects on trajectories running through that reference face.

**Method**

**Participants**

Sixteen Caucasian adult volunteers (three males) from the UWA community, aged 17-31 years ($M = 20.9$ years, $SD = 4.3$ years) participated. Participants either received course credit, or were reimbursed $20 for their time and travel expenses. One additional participant was tested but was excluded from
Stimuli

Stimuli were developed from those used by Skinner and Benton (2010). Their original stimuli were constructed from photographs of 25 male and 25 female faces posing the six basic expressions (happy, sad, angry, fearful, surprised and disgusted) and a neutral expression. For each expression the 50 images were combined using Psychomorph morphing software (Tiddeman, Burt, & Perrett, 2001) to create a single composite image of that expression. The composite of the neutral expressions created by this process was used as our neutral reference face. To produce an average reference face, the seven composite expressions were combined into a single composite image (Figure 2).

Figure 2. The neutral (left) and average (right) reference faces.

This participant performed normally for the anger, disgust and fear expressions at baseline, becoming more likely to correctly identify those expressions as expression intensity increased. However, the participant became less likely to correctly identify sadness as expression intensity increased on both the average-referenced and neutral-referenced trajectories at baseline.
Each expression trajectory was created by morphing an expression through one reference face and beyond it to an anti-expression. The features of the anti-expression differed from the reference face to the same extent as those of the original expression, but in the opposite direction. Expressions on these morph trajectories can be identified using percentages, where 100% indicates the original expression, 0% indicates the reference face and -100% indicates the anti-expression.

Because the neutral expression has a closed mouth, anti-expressions created by morphing along a trajectory from an open-mouthed expression through neutral contained some visual artifacts around the mouth. We avoided using the expressions that showed the worst of these artifacts (happiness and surprise), and chose anger, fear, disgust and sadness as our four target expressions. There were two trajectories for each expression, one made with the average reference face and one made with the neutral reference face, giving eight trajectories in total.

To meaningfully compare the size of the aftereffects produced by the neutral-referenced and average-referenced adaptors, the perceptual distance from the adaptors to the reference faces should be equal for both sets of trajectories (Rhodes & Jeffery, 2006), because more extreme adaptors produce stronger aftereffects (Burton et al., 2013; Skinner & Benton, 2010, 2012). To equate these distances, a separate group of participants rated the perceived similarity of the anti-expressions to their respective reference faces (see Supplementary Materials A). Participants viewed each anti-expression/reference face pair one at a time, and rated how similar the two faces looked on a seven-point scale. We used these ratings to select adaptor strengths for each trajectory such that for each expression the average-referenced anti-expression adaptor was judged to be as similar to the average reference face as the neutral-referenced anti-expression adaptor was to the
neutral reference face. The final similarity-matched adaptors are shown in Figure 3.

Test faces were taken from six levels along each of the eight trajectories (−20%, 0%, 20%, 40%, 60% and 80% expression strengths; Figure 4). Adaptors subtended a viewing angle of 8.8° by 11.9° when viewed from 50 cm. To reduce the effects of retinotopic low-level adaptation, test expressions were displayed at 75% of the size of the adaptors, subtending a viewing angle of 6.6° by 8.9°, and eye movements were allowed.
Table 3. First row shows the full-strength expression composites used to generate the anti-expressions. Next two rows are the full-strength anti-expressions made relative to the average and neutral anti-expressions. Bottom two rows are the final set of eight perceptually equated anti-expression adaptors. These adaptors were selected so that for each expression the average-referenced and neutral-referenced adaptors were matched for rated similarity to their respective reference faces.
Figure 4. Test faces. These were taken from eight trajectories (four expressions by two reference faces). For example, the average-referenced anger trajectory runs from anger, through the average reference face, to anti-anger. Test expressions were of six intensities, indicated on a percentage scale where 100% represents the full-strength expression, 0% indicates the reference expression and -100% represents the full-strength anti-expression.
Procedure

Testing was completed in two 45-minute sessions. Participants began each testing session by learning to identify the target expressions in a training task. They then completed a baseline block, in which they identified test expressions with no adaptation. Next, they were trained on a task in which they detected changes to the brightness of the eyes and lips of adaptors. We used this task to ensure that participants attended during adaptation, because better-attended adaptors result in larger aftereffects (Rhodes, Jeffery, et al., 2011). Finally, they completed an adaptation block in which they viewed an adaptor and then saw and identified a test expression on each trial.

Training. Participants were trained to identify the four target expressions (anger, disgust, fear and sadness), responding using marked keyboard keys. The aftereffect produces only a relatively weak impression of the target expression. By training participants to identify the target expressions we hoped to maximize their ability to categorize those weak expressions, and so maximize our ability to measure the aftereffect. Participants began by viewing the four target expressions (anger, disgust, fear and sadness) at 100% strength, one at a time, with an unlimited viewing duration. They identified the expressions using keyboard keys marked with the four target expressions. Participants viewed and identified the expressions one at a time in a series of blocks. Each block contained the four expressions, which were randomly reordered every time a block was presented. Participants continued until they identified two consecutive blocks with 100% accuracy (number of blocks to reach criterion ranged from 2-9, $M = 2.5, SD = 1.4$). Participants then moved on to a harder version of the training task, in which they identified the neutral- and average-referenced 60% versions of each expression. Participants again viewed and identified the expressions one at a time.
in a series of blocks with unlimited viewing duration. Each block contained the eight expressions, which were randomly reordered every time a block was presented. Participants continued until they identified two consecutive blocks with 100% accuracy (number of blocks to reach criterion ranged from 2-12, $M = 4.6$, $SD = 3.2$). Finally, participants were required to identify the same 60% expressions with viewing duration limited to 400ms. Again, they were required to identify two consecutive blocks of the eight expressions with 100% accuracy (number of blocks to reach criterion ranged from 2-8, $M = 3.4$, $SD = 2.0$).

Participants were also trained to identify the changes to the brightness of the eyes and lips that were used in the change detection task that took place during adaptation. They were first shown examples of the feature changes (Figure 5). They then saw a sequence of faces (the 0% neutral and average expressions) in 1000ms exposures, separated by 150ms blank ISIs. One quarter of these exposures had brightened lips, one quarter had brightened eyes, and the remaining faces had no brightened features. Participants used the marked keys to identify the changes they saw. They continued until they had responded correctly to a sequence of 16 consecutive trials, with the correct response being to press the correct key if a change was present, and to press neither key if no change was present.

**Baseline Block.** In each baseline trial participants viewed a test expression for 400ms, followed by a 150ms blank ISI. Participants were then shown a response screen, and identified the test expression using one of the four marked keys. The next trial began immediately. Every test face appeared six times in a baseline block (eight trajectories by six expression strengths by six repetitions = 288 total baseline trials per session). The trials were presented in random order. Participants took a short break half-way though the baseline block.
**Adaptation Block.** In each adaptation trial, an adaptor was shown for 4000ms, in four 1000ms exposures separated by 100ms blank ISIs. During adaptation, participants detected changes in the brightness of the eyes and lips of adaptors (a task designed to maintain participant attention). In each trial, either the second or third 1000ms adaptor exposure would contain either brightened irises or brightened lips (Figure 5). Participants identified which feature had changed by pressing one of two marked keys (“eyes” and “lips”) immediately upon seeing the change. The final exposure was followed by a 150ms blank ISI. The test expression was shown for 400ms, followed by another 150ms ISI and then a response screen. Participants identified the test expression using the marked keys. The next trial began immediately. The adaptor and test expression were always taken from the same trajectory (for instance, the 20% neutral-referenced anger test was always paired with the neutral-referenced anti-anger adaptor). Every test face appeared six times in each of the two testing sessions (eight trajectories by six expression strengths by six repetitions = 288 total adaptation trials per session). The trials were presented in a pseudo-random order designed to prevent a build-up of adaptation to any particular adaptor. The rules for this pseudorandom order were that all eight adaptors were required to appear in each set of eight consecutive trials, and that no adaptor was allowed to appear in two consecutive trials. Test strengths were distributed randomly throughout the block. Participants took a short break after every 36 trials. There were four practice trials demonstrating the full adaptation task (with participants both detecting feature changes and identifying the test face) before the beginning of the adaptation block.
Each session contained a baseline block and an adaptation block. Participants completed two testing sessions, resulting in a total of 576 baseline trials and 576 adapt trials.

Figure 5. Brightness changes used to maintain participant attention during adaptation: brightened irises (left) and lips (right).

**Results**

To measure the aftereffect, we plotted the proportion of test faces that were judged to show the target expression (the expression opposite the adaptor) as a function of test level. Adaptation should make participants more likely to identify the target expression at all test levels, shifting their response functions to the left from baseline to post-adapt. The size of this shift indicates the size of the aftereffect. We quantified this shift as the change in the 50% response threshold following adaptation.

Response rates were calculated individually for each participant. We calculated the proportion of target responses (e.g. ‘angry’ responses to faces on the neutral-referenced anger trajectory) for each test level on each of the eight
trajectories (four expressions by two reference faces), at baseline and after adaptation. In order to calculate the 50% response thresholds for these functions, we fit a mixed-effects logit model to the data for each expression (see Supplementary Materials C for details). Standard logistic regression fits a single set of parameters to all participants’ data. In our mixed-effects model the intercepts of the response functions were free to vary across participants to allow for consistent between-participant variability in responses. For each expression, we used a single model to simultaneously fit separate logistic functions for the average and neutral trajectories at baseline and post-adaptation. These functions are shown in Figure 6.

We began by testing whether each adaptor produced significant aftereffects. We calculated the 50% threshold for each function. The difference between the threshold at baseline and after adaptation represents the size of the aftereffect for each trajectory (Figure 7). We approximated the standard error of these aftereffects using the delta method (implemented by the nlcom function of the STATA data analysis package, StataCorp, 2013), allowing us to test the significance of these aftereffects. Aftereffects were significantly larger than zero for average-referenced anger ($z = 14.15, p < 0.001$), disgust ($z = 9.32, p < 0.001$), and sadness ($z = 7.07, p < 0.001$), and for neutral-referenced anger ($z = 10.98, p < 0.001$), disgust ($z = 9.54, p < 0.001$), and fear ($z = 8.44, p < 0.001$).

Aftereffects were not significantly larger than zero for average-referenced fear ($z = 1.50, p = .134$) and neutral-referenced sadness ($z = -1.17, p = .242$). Examination of Figure 6 shows that response rates were almost entirely above threshold at all levels in these conditions. It was therefore difficult to accurately estimate thresholds and so to detect any shift in the curve following adaptation, which may explain the lack of significant aftereffects in these conditions.
Figure 6. Functions fit by the mixed logit model to proportion of test faces judged to show the target expression (expression opposite the adaptor). Blue points show mean response rates at baseline; blue lines show associated baseline fit functions. Red points and lines show mean response rates and fit functions following adaptation. Dashed lines show the 50% response threshold. A leftward shift in the 50% response threshold following adaptation indicates an aftereffect.
Figure 7. Mean aftereffect size (measured as the difference in the 50% threshold between baseline and post-adapt) for each expression trajectory. A positive value indicates that adaptation made participants more likely to see the target expression (the expression opposite the adaptor). Error bars indicate standard error.

We compared the size of the aftereffect between the average and neutral trajectories for each expression using STATA’s nlcom function (StataCorp, 2013). For sadness, adaptation on the average trajectory produced a significantly larger aftereffect than adaptation on the neutral trajectory ($z = 3.42, p = .001, d = 1.35$). However, this difference may reflect the difficulty in measuring aftereffects for sadness on the neutral trajectory noted above. There was no significant difference in aftereffect size between average and neutral for the other three expressions: anger ($z = -0.58, p = .561, d = -0.21$), disgust ($z = 1.33, p = .183, d = 0.47$), or fear ($z = -1.67, p = .095, d = -0.64$).
Discussion

We adapted participants along trajectories that ran through either the average expression or a neutral expression. If one of these expressions is a better approximation of the norm of expression space than the other, then we should find larger aftereffects on the trajectories that run through that expression. We found no difference in aftereffect size between the average-referenced and neutral-referenced trajectories for anger, disgust or fear. For sadness we found significantly larger aftereffects on the average trajectory compared to the neutral trajectory, which could be taken to indicate that the average expression is a better approximation of the norm. However, this difference can likely be attributed to ceiling effects that limited our ability to detect aftereffects on the neutral-referenced sad trajectory but not on the average-referenced sad trajectory. Overall then, our results do not favour one reference face over the other as an approximation of the norm.

Although there are certainly differences between the average and neutral expressions (Figure 2), they are still relatively similar (and unexpressive) when compared to the wider range of emotions that can be conveyed by a face. It is therefore possible that both of these expressions lie near to the norm, and that this is the reason that we were unable to distinguish between them. Considering only the conditions in which there were no problems with curve fit (i.e. all but neutral-referenced sadness and average-referenced fear), we observe substantial expression aftereffects for both the average-referenced and neutral-referenced trajectories. Therefore, both reference faces may be reasonable approximations of the norm. It is also important to remember that the norm is likely to be slightly different from individual to individual, due to differences in experience, and that it will change within an individual over time. Therefore, it is likely that the
reference faces tested here were a better approximation of the norm for some participants than for others.

Several previous expression aftereffect studies have used the average expression we tested here as their approximation of the norm (Burton et al., 2015; Burton et al., 2013; Skinner & Benton, 2010, 2012), and others have used a neutral expression to approximate the norm (Juricevic & Webster, 2012; Rutherford et al., 2008; Rutherford et al., 2011). We were unable to demonstrate an advantage for one of these approximations over the other. However, this may not pose a problem for the interpretation of these previous studies. For many paradigms, it is sufficient that the “norm” is close enough to the true norm that it can be used to produce adaptors that selectively bias perception towards the target expressions. Additionally, the “norm” is often chosen as a test stimulus because it is ambiguous enough that adaptation produces measurable changes in the way it is perceived. We found here that both the average and neutral expressions functioned well as “norms” for the purpose of producing anti-expressions and expression aftereffects. The neutral expression does have the disadvantage of producing anti-expressions that have some visual artifacts around the mouth (see Figure 3, bottom row) that make these faces less naturalistic in appearance, but these anti-expressions still produced significant aftereffects.

The average expression that we tested here is only one of several possible “averages”. Our average expression was created by evenly weighting the six basic emotional expressions, and neutral, in a morph, as in many previous expression adaptation studies (Burton et al., 2015; Burton et al., 2013; Skinner & Benton, 2010, 2012). However, there are other ways that an “average norm” might be produced. The norm could reflect the mean of all previously seen expressions, weighted by frequency. Alternatively, more salient expressions could be weighted
more highly in the average. The norm might include only emotional expressions, as seen here, but it could also include other facial postures (e.g. speech, boredom, confusion). For instance, Cook et al. (2011) created their approximate norm by creating a morph average from video stills of an actress that included speech positions as well as emotional expressions. Even the wide range of expressions captured by this video does not necessarily capture the variation that is seen in day-to-day social interactions. One potentially fruitful approach (becoming increasingly achievable with advances in wearable technology) would be to record the faces that a person sees over an extended period, and produce an average expression that takes into account the true range and frequency of the expressions that that person encounters. If the norm represents the central tendency of a person’s experience of expressions, then an average tailored specifically to an individual’s experience may be the best approximation of the norm.

In conclusion, we found no evidence to suggest that either the average or the neutral expression is a better approximation of the norm. We suggest that both expressions may be reasonably close approximations of the norm, and that for practical purposes either may be useful when conducting expression adaptation research.
Acknowledgments

This research was supported by the Australian Research Council (ARC) Centre of Excellence in Cognition and its Disorders (CE110001021), an ARC Professorial Fellowship to GR (DP0877379), an ARC Discovery Outstanding Researcher Award to GR (DP130102300), and an Economic and Social Research Council – United Kingdom Grant to CPB (RES-000-22-4319). ALS was supported by the Medical Research Council (MC_UU_12013/7).
References


StataCorp. (2013). Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.


Supplementary Materials A: Matching adaptor similarity

To meaningfully compare the size of the aftereffects produced by the neutral-referenced and average-referenced adaptors, the perceptual distance from the adaptors to the reference faces should be equal for both sets of trajectories (Rhodes & Jeffery, 2006), because more extreme adaptors produce stronger aftereffects (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Skinner & Benton, 2010, 2012). To equate these distances, a separate group of participants rated the perceived similarity of the anti-expressions to their respective reference faces. Participants were fourteen Caucasian adults (four males) from the UWA community, aged 17-32 years (M = 24.3 years, SD = 4.4 years), who took part for course credit.

These participants were shown pairs of expressions: one anti-expression (strengths -10%, -20% … -100%) and the corresponding 0% reference face. Participants rated how similar the two faces looked on a scale from 1 (not at all similar) to 10 (very similar). We asked participants to rate similarity without specifying that we were interested in expression. We made this choice because during pilot testing we found that participants who were specifically asked to rate the similarity of expression had a tendency to rate faces that showed the same emotion as very similar, even when they differed in important visual attributes (e.g. feature positions, intensity of expression). Participants made each rating twice, once with the reference face on the left and once with it on the right. In total there were 160 trials (two repetitions by ten anti-expressions strengths by eight trajectories), which were presented in random order. Average ratings across participants are shown in Table S1.
Table S1: Mean ratings of the similarity of each anti-expression morph to its corresponding reference face (0% morph strength), on a scale from 1 (not at all similar) to 10 (very similar). Ratings were made for eight trajectories: four expressions (anger, disgust, fear and sadness) by two reference faces (average and neutral).

<table>
<thead>
<tr>
<th>Expression</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morph</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strength</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-10</td>
<td>8.6</td>
<td>9.3</td>
<td>8.8</td>
<td>8.9</td>
</tr>
<tr>
<td>-20</td>
<td>8.4</td>
<td>9.0</td>
<td>8.1</td>
<td>8.3</td>
</tr>
<tr>
<td>-30</td>
<td>8.3</td>
<td>8.7</td>
<td>8.0</td>
<td>7.8</td>
</tr>
<tr>
<td>-40</td>
<td>7.6</td>
<td>8.4</td>
<td>7.7</td>
<td>7.7</td>
</tr>
<tr>
<td>-50</td>
<td>7.3</td>
<td>7.9</td>
<td>6.4</td>
<td>6.5</td>
</tr>
<tr>
<td>-60</td>
<td>7.1</td>
<td>7.4</td>
<td>5.4</td>
<td>5.2</td>
</tr>
<tr>
<td>-70</td>
<td>7.0</td>
<td>6.8</td>
<td>5.1</td>
<td>5.0</td>
</tr>
<tr>
<td>-80</td>
<td>6.3</td>
<td>7.0</td>
<td>4.7</td>
<td>4.2</td>
</tr>
<tr>
<td>-90</td>
<td>6.0</td>
<td>6.0</td>
<td>4.2</td>
<td>3.9</td>
</tr>
<tr>
<td>-100</td>
<td>4.6</td>
<td>5.3</td>
<td>3.5</td>
<td>4.1</td>
</tr>
</tbody>
</table>

Our aim was to produce anti-expression adaptors for each expression that were well matched for perceived similarity to their reference faces. We began by looking at the ratings for the -100% anti-expressions (Table 1). As well as matching the perceived similarity of the adaptors, we also wanted to keep our anti-expression adaptors as extreme as possible to maximize the size of the aftereffects we produced. We achieved this by finding the trajectory that was judged as the less extreme of the two for each expression (higher similarity ratings at -100%), and setting the -100% anti-expression as the adaptor for this trajectory. For example, for anger the neutral-referenced trajectory is less extreme than the
average-referenced trajectory, with a higher similarity rating, so for neutral-referenced anger we chose -100% anti-anger as the adaptor. We then found the morph level on the second trajectory that was matched to the first adaptor in perceptual similarity. In the case of anger, this means finding an average-referenced anti-anger morph level with a similarity rating of 5.3, to match -100% neutral-referenced anti-anger. To find this level, we fit Gaussian functions to the similarity ratings on the second trajectory. We then found the point on the curve where the similarity rating matched that of the first adaptor. This procedure gave a set of 8 anti-expression adaptors with the same mean perceived similarity to the reference face for average and neutral trajectories (Figure 3 of main text).
Supplementary Materials B: The mixed logit model

The adaptation aftereffect produces a change in participants’ responses from baseline to post-adaptation, such that their likelihood of responding with the expression opposite the adaptor (the ‘target’ expression) increases following adaptation. To estimate the size of the aftereffect, we can fit a curve to participants’ likelihoods of responding with the target expression across the various test levels both before and after adaptation, and measure the extent to which the 50% response threshold shifts with adaptation.

We modeled participants’ response data using a mixed logit model (see Burton et al. 2015 for a more detailed description of this model). In this analysis we are concerned with only two alternatives (whether a participant did or did not respond with the target expression). In its standard form for a two-alternative choice, the logit model gives the probability of making choice $i$ as:

$$P_{ni} = \frac{e^{\beta_i x_n}}{1 + e^{\beta_i x_n}}$$

Where participant $n$ makes choice $i$ from choice set $j$. The link function ($\beta_i x_n$) can be expanded to:

$$\beta_{ni} x = b_i x + a_{ni}$$

In our case, $x$ is the expression strength of the test face, and $b$ and $a$ determine the slope and intercept of the response function. We were interested in comparing the size of the aftereffects between the average-referenced trajectory and the neutral-referenced trajectory for each expression. To implement this comparison using
STATA statistics software, we fit the response functions for all four relevant conditions (baseline and post-adaptation for each of the average- and neutral-referenced trajectories) within a single model for each expression. For ease of description, we will refer to these conditions as “A0” (average-referenced baseline), “A1” (average-referenced post-adapt), “N0” (neutral-referenced baseline) and “N1” (neutral-referenced post-adapt). We will describe the full model below – however, it should be noted that fitting the four functions together in this way is equivalent to fitting each function in a separate model.

The response function for condition A0 is described by the parameters $b_i$ and $a_{ni}$ in the basic link function described above. For each of the remaining conditions, two additional parameters were added. These parameters adjust the slope and intercept of the first function in order to describe the other functions:

$$
\beta_{ni}^{A1} x = (b_i + bA1_i)x + a_{ni} + aA1_i
$$

$$
\beta_{ni}^{N0} x = (b_i + bN0_i)x + a_{ni} + aN0_i
$$

$$
\beta_{ni}^{N1} x = (b_i + bN1_i)x + a_{ni} + aN1_i
$$

The standard formulation of the logit model assumes that the parameters $\beta_i$ that describe the shape of the probability function are the same across all participants. In our case, it is likely that these parameters in fact vary systematically between individuals. For instance, one participant might have a consistent bias to respond “angry” more frequently than other participants, such that their baseline response curve for that expression is shifted closer to zero relative to other participants. The mixed logit model allows for these systematic differences between participants by allowing some or all parameters to be random: that is, to take different values for different participants. Rather than
estimating a single parameter value, for a random parameter we estimate the
distribution of the values that it takes across our sample of participants.

In our model, we made the intercept term $a_{oi}$ a random parameter. This is
equivalent to having a fixed constant and an additional random error with mean
zero that captures consistent differences between participants. However, the effect
of participant will not necessarily be consistent across all conditions. For instance,
it is possible that a participant may be more likely than other participants to
respond “anger” for the average-referenced test images, but less likely than other
participants to respond “anger” for the neutral-referenced test images. For this
reason, we set our random parameter to vary at the level of participant*reference
face. The random effect for a given participant is the same across the functions for
A0 and A1 (average-referenced trajectory before and after adaptation), and across
the functions for N0 and N1 (neutral-referenced trajectory before and after
adaptation), but is allowed to vary between the two sets of functions.

The analysis was run using the GSEM function of the Stata statistics
package (Statacorp, 2013). The estimated parameters for each of the four
expressions are given in Tables S2-5. The functions described by these parameters
are graphed in Figure 6 in the main text.
Table S2. Parameters estimated by the mixed logit analysis for Anger. For the random parameter we estimate the mean and variance of the random distribution.

Log Likelihood = -1511.69.

|      | ANGER | Coef. | Std. Err | z     | P>|z| | 95% Conf. Interval |
|------|-------|-------|----------|-------|------|-------------------|
| Fixed Parameters |
| b    | 0.08  | 0.00  | 18.46    | 0.000 | 0.07 | 0.07 - 0.09       |
| bA1  | 0.01  | 0.01  | 1.83     | 0.068 | 0.00 | 0.00 - 0.02       |
| bN0  | -0.02 | 0.01  | -3.00    | 0.003 | 0.03 | -0.03 - 0.01      |
| bN1  | -0.01 | 0.01  | -1.08    | 0.280 | 0.02 | -0.02 - 0.01      |
| aA1  | 1.66  | 0.22  | 7.52     | 0.000 | 1.23 | 1.23 - 2.09       |
| aN0  | 2.16  | 0.44  | 4.96     | 0.000 | 1.31 | 1.31 - 3.01       |
| aN1  | 3.82  | 0.43  | 8.79     | 0.000 | 2.97 | 2.97 - 4.67       |
| Random Parameter |
| Mean a| -3.02 | 0.33  | -9.27    | 0.000 | -3.65| -3.65 - -2.38     |
| Variance Estimates for Random Parameter |
| $\sigma^2$ | 1.15  | 0.32  | 0.67     | 1.97  |
Table S3. Parameters estimated by the mixed logit analysis for Disgust. For the random parameter we estimate the mean and variance of the random distribution.

Log Likelihood = -1311.13.

| DISGUST | Coef. | Std. Err | z    | P>|z| | 95% Conf. Interval |
|---------|-------|----------|------|------|-------------------|
| **Fixed Parameters** |       |          |      |      |                   |
| $b$      | 0.07  | 0.00     | 18.26| 0.000| 0.07              | 0.08              |
| $bA1$    | 0.02  | 0.01     | 3.01 | 0.003| 0.01              | 0.03              |
| $bN0$    | 0.04  | 0.01     | 4.67 | 0.000| 0.02              | 0.05              |
| $bN1$    | 0.04  | 0.01     | 4.96 | 0.000| 0.02              | 0.05              |
| $aA1$    | 1.17  | 0.16     | 7.49 | 0.000| 0.86              | 1.47              |
| $aN0$    | -2.96 | 0.39     | -7.51| 0.000| -3.73             | -2.19             |
| $aN1$    | -1.61 | 0.35     | -4.56| 0.000| -2.30             | -0.92             |
| **Random Parameter** |       |          |      |      |                   |
| Mean $a$ | -1.25 | 0.22     | -5.63| 0.000| -1.69             | -0.82             |
| **Variance Estimates for Random Parameter** |       |          |      |      |                   |
| $\sigma^2$ | 0.57  | 0.17     | 0.32 | 1.02 |                   |
Table S4. Parameters estimated by the mixed logit analysis for Fear. For the random parameter we estimate the mean and variance of the random distribution.

Log Likelihood = -1286.06.

|                | Coef. | Std. Err | z     | P>|z|  | 95% Conf. Interval |
|----------------|-------|----------|-------|------|-------------------|
| **FEAR**       |       |          |       |      |                   |
| **Fixed Parameters** |       |          |       |      |                   |
| $b$            | 0.04  | 0.00     | 13.36 | 0.000| 0.04              |
| $bA1$          | 0.04  | 0.01     | 4.86  | 0.000| 0.02              |
| $bN0$          | 0.06  | 0.01     | 9.15  | 0.000| 0.05              |
| $bN1$          | 0.10  | 0.01     | 10.28 | 0.000| 0.08              |
| $aA1$          | 0.89  | 0.15     | 6.18  | 0.000| 0.61              |
| $aN0$          | -4.07 | 0.32     | -12.64| 0.000| -4.71             |
| $aN1$          | -3.82 | 0.34     | -11.38| 0.000| -4.48             |
| **Random Parameter** |       |          |       |      |                   |
| Mean $a$       | 0.58  | 0.17     | 3.36  | 0.001| 0.24              |
| **Variance Estimates for Random Parameter** |       |          |       |      |                   |
| $\sigma^2$     | 0.36  | 0.12     |       |      | 0.19              |
Table S5. Parameters estimated by the mixed logit analysis for Sadness. For the random parameter we estimate the mean and variance of the random distribution.

Log Likelihood = -1645.48.

|          | Coef. | Std. Err | z    | P>|z| | 95% Conf. Interval |
|----------|-------|-----------|------|------|-------------------|
| **SADNESS** |       |           |      |      |                   |
| **Fixed Parameters** |       |           |      |      |                   |
| b        | 0.07  | 0.00      | 18.63| 0.00 | 0.06              |
| bA1      | 0.02  | 0.01      | 2.82 | 0.01 | 0.01              |
| bN0      | -0.03 | 0.01      | -6.79| 0.00 | -0.04             |
| bN1      | -0.01 | 0.01      | -2.42| 0.02 | -0.02             |
| aA1      | 0.56  | 0.20      | 2.82 | 0.01 | 0.17              |
| aN0      | 3.33  | 0.49      | 6.80 | 0.00 | 2.37              |
| aN1      | 3.40  | 0.49      | 6.95 | 0.00 | 2.44              |
| **Random Parameter** |       |           |      |      |                   |
| Mean a   | -2.51 | 0.36      | -7.06| 0.00 | -3.20             |
| **Variance Estimates for Random Parameter** |       |           |      |      |                   |
| $\sigma^2$ | 1.64  | 0.45      |      | 0.96 | 2.80              |
References


Chapter Four

The Timecourse of Expression Aftereffects
Abstract

Adaptation to facial expressions produces aftereffects that bias perception of subsequent expressions away from the adaptor. Studying the temporal dynamics of an aftereffect can help us to understand the neural processes that underlie perception, and how they change with experience. Little is known about the temporal dynamics of the expression aftereffect. We conducted two experiments to measure the timecourse of this aftereffect. In Experiment 1 we examined how the size of the aftereffect varies with changes in the duration of the adaptor and test stimuli. We found that the expression aftereffect follows the classic timecourse pattern of logarithmic build-up and exponential decay that has been demonstrated for many lower-level aftereffects, as well as for facial identity and figural face aftereffects. This classic timecourse pattern suggests that facial expression is processed in similar ways to lower-level visual stimuli, and that the expression aftereffect is likely to be perceptual in origin. We also found that aftereffects could be generated by as little as 1 second of adaptation, and in some conditions lasted for as long as 3200 ms. We extended this last finding in Experiment 2, exploring the longevity of the expression aftereffect by adding a stimulus-free gap of varying duration between adaptation and test. We found that significant expression aftereffects were still present 32 seconds after adaptation. The persistence of the expression aftereffect suggests that they may have a considerable impact on day-to-day expression perception.
The Timecourse of Expression Aftereffects

A fascinating property of the visual system is that many aspects of visual perception are adaptive, such that the appearance of a given stimulus may be affected by what has been seen before. For instance, after viewing a waterfall for a period of time the visual system adapts to that downward motion and a stationary surface viewed immediately afterwards will temporarily appear to be moving upwards (Anstis, Verstraten, & Mather, 1998). This perceptual bias is known as an aftereffect.

Although many of the documented adaptation aftereffects concern lower-level visual properties like motion, adaptation also occurs for higher-level stimuli, such as faces. In the facial expression aftereffect, adaptation to a face with a particular expression will bias participants’ judgments of subsequent faces towards the ‘opposite’ expression: the expression with visual characteristics opposite those of the adaptor, relative to the central tendency of expressions. For example, Figure 1 shows fear and its ‘opposite’ expression, anti-fear. Where fear has raised eyebrows and an open mouth, anti-fear has lowered eyebrows and a closed mouth, and so on. Adapting to anti-fear produces a selective perceptual bias towards fear (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Skinner & Benton, 2010, 2012). Adaptation aftereffects can also be observed for other facial attributes, including identity (Leopold, O'Toole, Vetter, & Blanz, 2001), gender and race (Webster, Kaping, Mizokami, & Duhamel, 2004), and gaze direction (Jenkins, Beaver, & Calder, 2006).
Figure 1. Fear (left) and anti-fear (right). The feature positions of these two expressions are opposite one another relative to the central tendency of facial expressions.

Aftereffects are particularly useful to researchers because they reflect changes in the responsiveness of the neurons that code the adapted property, and so can be used to reveal the nature of the coding mechanisms that underlie visual perception. One fruitful avenue of study is the temporal dynamics of aftereffects. By studying the timecourse of an aftereffect, researchers can describe how quickly the visual system adapts to current stimuli, how the length of adaptation affects the strength of the aftereffect, and how long the adapted state persists. There has been some exploration of the timecourses of face-related adaptation (e.g. facial identity: Leopold, Rhodes, Müller, & Jeffery, 2005; Rhodes, Jeffery, Clifford, & Leopold, 2007; distortion of familiar faces: Carbon & Ditye, 2011, 2012; Strobach, Ditye, & Carbon, 2011; gaze direction: Kloth & Schweinberger, 2008). However, little is known about the timecourse of the expression aftereffect.

We conducted two experiments to measure the temporal dynamics of expression adaptation. Our first aim was to determine whether the expression
aftereffect follows the “classic” timecourse pattern of logarithmic increase and exponential decay found for many lower-level visual aftereffects. Face aftereffects are different to lower-level visual aftereffects in many respects. The stimuli themselves are more complex, and the aftereffects persist over changes in the size, position and angle of stimuli between adaptation and test (Leopold et al., 2001; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Watson & Clifford, 2003; Zhao & Chubb, 2001). However, this same classic timecourse pattern has been shown for both facial identity and figural face aftereffects (Leopold et al., 2005; Rhodes et al., 2007), and is not simply inherited from lower-level adaptation because it remains even with a size change between adaptation and test (Rhodes et al., 2007). This shared timecourse pattern suggests that these face aftereffects, like lower-level effects, are perceptual in nature, and do not simply reflect changes in participants’ response strategies. We aimed to determine whether the same is true for expression aftereffects.

In our first experiment, participants adapted to anti-expressions for varying amounts of time, and then viewed an ambiguous test expression for varying amounts of time. They then rated the strength of the expression they perceived on the test face. This design allowed us to measure how aftereffect strength varied with adapt duration (the build-up of the aftereffect) and with test duration (the decay of the aftereffect).

The design of Experiment 1 also allowed us to observe which of our conditions produced expression aftereffects - whether our shortest adaptation durations were long enough to result in an aftereffect, and whether the aftereffects lasted long enough to be measured at the end of our longest test durations. To foreshadow our results from this first experiment, we found that expression aftereffects were still present after our longest test duration (3200 ms). We
extended this finding in Experiment 2 by measuring how long the expression aftereffect persists. An interesting feature of visual adaptation is that it can persist for a surprisingly long time. Following 15 minutes of adaptation, a motion aftereffect can in some conditions be observed 24 hours later. Fifteen minutes of adaptation can produce a McCollough effect that lasts as long as three months (Jones & Holding, 1975). Considerable longevity has also been found for some face aftereffects: for instance, adapting to a distorted famous face for 25 minutes can result in altered perception of that identity as much as a week later (Carbon & Ditye, 2011, 2012). A dynamic face stimulus like facial expression might be less likely to produce such long-lasting aftereffects. However, for gaze direction, another rapidly-changing dynamic face cue, 84 seconds of adaptation can produce aftereffects that last for seven minutes (Kloth & Schweinberger, 2008); certainly long enough to cover many changes in gaze direction during conversation. We were interested in the longevity of the expression aftereffects created by the relatively brief adaptation durations (up to 16 seconds) used in Experiment 1. These seconds-long adaptation durations are likely to be representative of the duration of expression adaptation that might be experienced in a naturalistic setting.

**Experiment 1**

The aim of Experiment 1 was to map the effects of both adaptation and test durations on aftereffect size. In particular, we aimed to determine whether expression aftereffects show the classic timecourse found for both lower-level visual aftereffects and facial identity and figural face aftereffects. On each trial, participants adapted to an anti-expression and then rated the strength of the opposing expression in an ambiguous test face. We used static expressions as adaptors and test faces, which allowed us to straightforwardly vary the duration of
the adapt and test exposures. We expected that, following the classic timecourse, aftereffect magnitude would increase logarithmically with increased adaptation time and decrease exponentially with increased test time.

Method

Participants. Twelve adults (5 male) participated. Mean age was 23.8 years, SD = 6.1 years. Participants were either first year psychology students participating for course credit, or were compensated $15 for their travel costs.

Stimuli. Stimuli were adapted from Skinner and Benton’s (2010) gender-neutral expressive faces created from images of 50 Caucasian individuals (25 male and 25 female) posing various expressions. For each expression the 50 images were morphed together using Psychomorph (Tiddeman, Burt, & Perrett, 2001) to create a single identity-neutral image that captured the key characteristics of that expression while minimizing any individual idiosyncrasies. The test face was an average expression created by morphing together seven of these identity-neutral expressions (anger, disgust, fear, happiness, sadness, surprise and neutral) (Figure 2).

Figure 2. The average expression, created by morphing together seven expressions (anger, disgust, fear, happiness, sadness, surprise and neutral) derived from 20 identities.
The adaptors were anti-expressions, each created by morphing along a trajectory that ran from one of the identity-neutral expressions, through the average expression and beyond it to a point that differed from the average to the same extent as the original expression. Six anti-expression adaptors (anti-anger, anti-disgust, anti-fear, anti-happiness, anti-surprise and anti-sadness) were created in this way (Figure 3). Points on each trajectory are labeled as percentages: “100%” for the original expression, “0%” for the average expression and “-100%” for the anti-expression. During training, we used weaker versions of the original expressions taken from these trajectories (e.g. 50% fear, which lies halfway between the average expression and fear).

Figure 3. Anti-expression adaptors.
**Procedure.** Participants began by learning to identify the six target expressions (anger, disgust, fear, happiness, sadness and surprise), responding using marked keys. We expected that after adaptation, the average test expression would be perceived as a weak version of the target expression. By training participants to identify the target expressions we hoped to maximize their ability to categorize those weak expressions, and so maximize our ability to measure the aftereffect.

Each expression was shown for 400ms. Participants identified the expressions in two stages, first at 90% strength and then at 50% strength. At each stage they were only able to move on once they had correctly identified a sequence of all six expressions twice consecutively. Participants required $M = 3.8$ (SD = 2.7) repetitions of the six-expression sequence to meet criterion for the 90% strength expressions, and $M = 4.9$ (SD = 4.7) repetitions to meet criterion for the 50% expressions.

Participants then learned to rate the strength of expressions. Each rating trial began with a cue word (X) that gave the expression that participants would be rating in that trial. Participants then saw a face for 400ms. Finally participants were asked “how strong was your impression of (X)?” which they answered on a 7-point scale from 1, “No (X)”, to 7, “Strong (X)” using labelled keyboard keys. Participants rated each of the six target expressions at six strengths: 90%, 70%, 50%, 30%, 10% and 0% (average). These expressions were presented in random order. Before making their ratings, participants saw sequence of faces that demonstrated the range of variation in the set (19 images: 10%, 50% and 90% versions of each expression, plus the average). Participants repeated the rating training if their ratings demonstrated that they did not understand the task. Experimenters checked that participants used higher ratings for more intense
expressions, and that participants attempted to use the full range of the scale. Only one participant was required to repeat the rating training, and only repeated it once.

The adaptation procedure followed Leopold et al. (2005) and Rhodes et al. (2007). Each trial began with a cue to the expression that participants would be rating, as above. Participants then viewed an anti-expression adaptor, followed by a test face (always the average expression), and rated the strength of their impression of the cued expression using the same scale that was used in the training task. Because impressions could be dynamic, participants were asked to rate their impression at the offset of the test face.

Timings were as follows: a fixation cross (200ms), then the expression cue (1000ms), an adaptor (variable duration), an ISI (150ms), the test face (variable duration), and then the response screen with rating scale. A beep sounded 250ms before the end of the adapting face to warn participants that the test face was about to appear. Following Leopold et al. (2005) and Rhodes et al. (2007), there were five adaptation durations (1000, 2000, 4000, 8000, 16000ms) and five test durations (200, 400, 800, 1600, 3200ms). Each combination of adaptation and test duration was used in 12 trials, two with each of the six expressions (300 trials total). Trials were divided into 10 blocks of 30, containing an equal number of trials with each expression and each adapt and test duration (not fully crossed). Trials were randomized within blocks. All trials were completed in a single session, with participant-timed breaks between blocks. The rating task began with twelve practice trials; in these trials the adaptor was the average expression (shown for 1000ms, 2000ms or 4000ms to introduce participants to the variable nature of the adaptor durations) and the test face was a 30% version of one of the
six target expressions, shown once for 200ms and once for 1600ms. The entire session, including training and adaptation, took around 90 mins to complete.

**Results and Discussion**

The strength of the aftereffect was measured as the strength of the impression of a target expression following adaptation to its anti-expression. For each participant an average aftereffect size was found for each combination of adapting duration and test duration, collapsed across the six target expressions (Figure 4).

We began by examining what effect adaptation duration and test duration had on the size of the aftereffect. Ratings of expression intensity increased with increasing adapting duration (adapt 1 s: $M = 3.55, SD = 0.89$; adapt 2 s: $M = 3.72, SD = 0.98$; adapt 4 s: $M = 3.82, SD = 0.99$; adapt 8 s: $M = 3.99, SD = 1.03$; adapt 16 s: $M = 4.27, SD = 1.0$) and decreased with increasing test duration (test 200 ms: $M = 4.34, SD = 1.05$; test 400 ms: $M = 3.96, SD = 0.98$; test 800 ms: $M = 3.81, SD = 0.99$; test 1600 ms: $M = 3.68, SD = 0.96$; test 3200 ms: $M = 3.57, SD = 1.01$). We used a two-way repeated measures ANOVA with a Greenhouse-Geisser correction to test these effects. There was a significant main effect of adapt duration, $F(1.478, 16.254) = 23.30, p < .001, \eta_p^2 = .679$ and a significant main effect of test duration, $F(1.487, 16.352) = 18.60, p < .001, \eta_p^2 = .628$, with no significant interaction, $F(6.527, 71.796) = 1.49, p = .189, \eta_p^2 = .119$. These results confirm that both adaptation duration and test duration had significant effects on the size of the expression aftereffect.
Figure 4. Mean size of expression aftereffects as a function of adapt duration and test duration (N=12). Size of aftereffect was measured as the strength of impression of the cued expression.

To examine whether the data showed the expected pattern of logarithmic build-up and exponential decay, we plotted the ratings at each adaptation duration (collapsed across test duration) and at each test duration (collapsed across adaptation duration) on semi-log coordinates (Figure 5, C-D; also shown on untransformed coordinates, A-B). We used relative ratings, calculated by subtracting each participant’s grand mean from their ratings (Leopold et al., 2005;
Rhodes et al., 2007). This adjustment accounts for any overall biases in participants’ responses (e.g. a tendency to rate all expressions lower on the scale than other participants), allowing us to more clearly see any patterns across participants’ responses. If the data follow the expected pattern the points should form straight lines when plotted on semi-log coordinates. Straight line fits to the group data (Figure 5, C-D) were excellent, with $R^2 = 0.97$ for the adaptation duration function (slope = 0.57) and $R^2 = 0.92$ for the test duration function (slope = -.61). Runs tests indicated no significant non-linearities ($p > 0.50$). These results confirm that expression aftereffects follow the classic timecourse pattern of logarithmic build-up and exponential decay found for lower-level visual aftereffects. This pattern is also consistent with the timecourses found for other face aftereffects (Leopold et al., 2005; Rhodes et al., 2007).
Figure 5. Relative ratings (means ± SE) as a function of adapt and test duration. Relative ratings calculated by subtracting each participant’s grand mean from their ratings. (a) Relative ratings as a function of adapting time. (b) Relative ratings as a function of test duration. (c) Relative ratings plotted on semi-log coordinates, as a function of adapting time, with line of best fit. (d) Relative ratings plotted on semi-log coordinates, as a function of test duration, with line of best fit.

A secondary aim was to determine which adaptation durations produced significant aftereffects, and whether aftereffects were still present after our longest
test durations. Single-sample t-tests confirmed that ratings were significantly higher than zero in all conditions (all ps < .001). However, it is unlikely that expression ratings would fall to zero in the absence of an aftereffect because the average expression contains each of the target expressions and may resemble each of them to some extent. Our closest approximation of a no-adaptation rating is in the condition where the aftereffect is expected to be at its smallest: 1 second adapt duration and 3200ms test duration (Adapt1Test3200). In this condition there should be either no aftereffect remaining, or a small aftereffect. Significantly higher ratings in other conditions can therefore be taken as evidence of an aftereffect in those conditions.

Inspection of Figure 4 suggests that aftereffects were produced after as little as 1 sec of adaptation: when test durations were low (e.g. 200ms test), ratings following 1 sec adaptation appear higher than our approximate baseline Adapt1Test3200. A paired samples t-test confirmed that there was a significant aftereffect in this condition, \( t(11) = 4.14, p = .002, d = 0.63 \) (bonferroni-corrected alpha = .01), indicating that the expression aftereffect can be generated by as little as one second of adaptation.

It also appears that aftereffects remained for as long as 3200ms of test exposure, at least in the longer adaptation conditions. To determine what length of adaptation was required to produce a significant aftereffect that remained after 3200ms of test exposure, we used paired samples t-tests to compare ratings of expression intensity at Adapt1Test3200 to ratings at Adapt2Test3200, Adapt4Test3200, Adapt8Test3200 and Adapt16Test3200. There was a significant aftereffect after 16 seconds of adaptation, \( t(11) = -5.03, p < .001, d = 0.54 \), and after 8 seconds of adaptation, \( t(11) = -3.12, p = .010, d = 0.40 \). There was no significant aftereffect after 4 seconds of adaptation, \( t(11) = -2.33, p = .040, d = \)
0.29, or 1 second of adaptation, \( t(11) = -1.23, p = .245, d = 0.16 \) (with bonferroni-corrected alpha = .01). These results show that adaptation of a sufficiently long duration (at most 8 s, and possibly 4 s) can produce an aftereffect that remains after 3200 ms of exposure to the test face.

**Experiment 2**

In Experiment 1 we saw significant aftereffects that lasted for at least 3200 ms. Here we tested how much longer the aftereffect might persist. We modified the method of Experiment 1, using only the 200ms test duration and inserting a stimulus-free gap of varying duration between adaptation and test. In Experiment 1, the duration of the test face varied, and participants rated the expression they saw immediately before offset of the test face. In this experiment, we used a fixed test duration and a gap between adapt and test to more explicitly test the persistence of the aftereffect. We were concerned that if we just extended the duration of the test face, as in Experiment 1, participants might anchor their ratings to the stronger aftereffects experienced earlier in the test exposure (see Figure 4), causing us to overestimate the persistence of the aftereffect. We chose an adaptation duration of 16 seconds because it produced the largest aftereffects in Experiment 1. We also included an eight second adaptation condition. This duration also produced large aftereffects in Experiment 1 (including a significant aftereffect after 3200 ms of exposure to the test face). This shorter adaptation duration might more closely resemble the durations of expressions seen in typical interactions than the longer 16 second exposures. For instance, spontaneous expressions elicited by film clips have been found to range from an average duration of 3 seconds (Frank, Ekman, & Friesen, 1993) to 13 seconds (Pfister, Li, Zhao, & Pietikainen, 2011). Adapt-test gap durations were chosen from both
within and beyond the range of test periods from Experiment 1 (500 ms, 1000 ms, 4000 ms, 32000 ms).

In Experiment 1, adaptation durations varied sufficiently to hold participant attention throughout the adaptation period (as the test stimulus could appear at any time). In this experiment both adaptation durations were relatively long. Therefore, to maintain participant attention during adaptation, we included a change detection task in which participants identified changes in the brightness of the eyes or lips of the adaptors (Burton et al., 2013). In this experiment we also included a baseline rating measure for each expression in the absence of adaptation. As discussed above, even if no adaptation occurs participants are unlikely to rate the intensity of any given target expression as zero, because the test face is made up of a combination of the target expressions. Comparison of post-adaptation ratings to baseline ratings provides an explicit test for the aftereffect.

Method

Participants. Twenty-three adults (11 male) participated in the study (24 participants were tested, but one was excluded from analysis: see results). Mean age was 19.6 years, SD = 2.7 years. Participants were first year psychology students participating for course credit. Our sample was larger than in Experiment 1 because we expected that aftereffects would be reduced given the longer delays between adapt and test images.

Stimuli. Stimuli were the same as those used in Experiment 1. For the change detection task that took place during adaptation, versions of the adaptors were created which had slightly brightened irises or lips. These were made by overlaying a white mask onto these features at 10% opacity (Figure 6).
Procedure. Participants began with the same training tasks as in Experiment 1. Next, we found each participant’s baseline expression intensity rating for the average test face, for each expression. Participants saw a fixation cross (200ms), then an expression cue (1000ms), the test face (always the average expression, 200ms), then the response screen with rating scale. Participants rated the perceived intensity of each of the six target expressions ten times, giving 60 trials total.

Each trial of the adaptation task began with a cue to the expression that participants would be rating, as in Experiment 1. Participants then viewed an anti-expression adaptor. There was then a delay period, followed by a test face (always the average expression). Participants rated the strength of their impression of the cued expression.

Timings were as follows: A fixation cross (200ms), then the expression cue (1000ms), an adaptor (either 8 or 16s), a delay period (500 ms, 1000 ms, 4000 ms, 32000 ms), the test face (200ms) and then the response screen with rating.
scale. A 200ms beep sounded 400ms before the end of the delay period to warn participants that the test face was about to appear. Participants were given no special instructions about what to do during the delay. To ensure that they were not exposed to any other faces during this time testing was conducted in an otherwise empty room, with the experimenter seated out of view behind the participant. An anti-glare film was applied to the monitor to prevent participants from seeing their own reflection.

To maintain attention during adaptation, participants were asked to detect temporary changes in the brightness of the irises or lips of the adaptors. To facilitate this change detection task, adaptors were shown over repeated 1000-ms exposures, separated by 150-ms ISIs. During each adaptation period, one of these 1000-ms exposures included either an eye or a lip change. Participants responded using marked keyboard keys as soon as they saw a change.

Each combination of target expression, adaptation duration and delay duration (six expressions by two adaptation durations by four delay durations = 48 combinations) was used in six trials, for a total of 288 trials. Trials were presented in two sessions, each containing an identical set of 144 trials. Trials were presented in a pseudo-random order, in which each run of 6 trials must contain each of the 6 expressions, and no two consecutive trials could contain the same expression. Participants took breaks every 24 trials.

The baseline block was repeated at the end of each testing session to confirm that baseline ratings had not shifted during the session. In total, each session took approximately 90 minutes for a total of 3 hrs of testing.

**Results and Discussion**

We examined participants’ eye and lip change performance during adaptation to check that they had been attending during the task. One participant
(from the original sample of 24) was removed from analysis because of poor change detection performance in session two (27% eye changes and 14% lip changes correctly detected). The remaining participants correctly detected a mean of 92.8% of eye changes (SD = 7.7%) and 87.0% of lip changes (SD = 12.3%) in session 1, and 90.4% of eye changes (SD = 10.3%) and 88.5% of lip changes (SD = 11.9%) in session 2.

There was no significant difference between baseline ratings at the start and end of the adaptation task, or between sessions (see supplementary materials). We therefore took an average of the ratings from all baseline blocks for each subject to give the final baseline rating for each expression.

To calculate each participant’s aftereffects, we calculated their average rating for each combination of expression, adapt duration and gap duration across the two sessions, and subtracted their average baseline rating for that expression. We then took an average across all expressions to calculate each participant’s aftereffect for each combination of adaptation duration and gap duration (Figure 7).

![Figure 7](image-url)  
*Figure 7*. Mean aftereffect size (post-adapt ratings minus baseline ratings) for each adaptation and gap duration. Bars show standard error.
To determine whether all conditions produced significant aftereffects, we compared each aftereffect score to zero using one-sample t-tests. All aftereffects were significantly larger than zero (Table 1).

Table 1: One-Sample T-Tests Comparing Aftereffect Score to Zero in Each Condition

<table>
<thead>
<tr>
<th>Gap Duration</th>
<th>Adapt 8</th>
<th></th>
<th></th>
<th>Adapt 16</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t(22)</td>
<td>p</td>
<td>d</td>
<td>t(22)</td>
<td>p</td>
<td>d</td>
</tr>
<tr>
<td>500 ms</td>
<td>9.00</td>
<td>&lt; .001</td>
<td>1.88</td>
<td>11.01</td>
<td>&lt; .001</td>
<td>2.30</td>
</tr>
<tr>
<td>1000 ms</td>
<td>8.30</td>
<td>&lt; .001</td>
<td>1.73</td>
<td>10.68</td>
<td>&lt; .001</td>
<td>2.23</td>
</tr>
<tr>
<td>4000 ms</td>
<td>9.09</td>
<td>&lt; .001</td>
<td>1.89</td>
<td>12.55</td>
<td>&lt; .001</td>
<td>2.62</td>
</tr>
<tr>
<td>32000 ms</td>
<td>7.45</td>
<td>&lt; .001</td>
<td>1.55</td>
<td>8.73</td>
<td>&lt; .001</td>
<td>1.82</td>
</tr>
</tbody>
</table>

We tested the effects of adaptation duration and gap duration on aftereffect size using a within-subjects ANOVA. We found a significant main effect of adaptation duration, $F(1, 22) = 49.87, p < .001, \eta_p^2 = .694$, a significant main effect of gap duration, $F(3, 66) = 27.75, p < .001, \eta_p^2 = .558$, and no significant interaction, $F(3, 66) = 1.94, p = .131, \eta_p^2 = .081$. As expected, aftereffects were larger for 16 sec adaptation than for 8 sec adaptation, and decreased in size as gap duration increased. However, as noted above, aftereffects were still present after an adapt-test gap of 32 seconds for both 16 sec and 8 sec of adaptation. This finding shows that expression aftereffects that could reasonably be generated during typical social interactions can persist for a surprisingly long time, given the fast-changing nature of expressions.
General Discussion

In Experiment 1, we found that the expression aftereffect followed the classic timecourse of logarithmic build up and exponential decay found for lower-level visual aftereffects. Similar timecourses have been found for facial identity and figural face aftereffects (Leopold et al., 2005; Rhodes et al., 2007). Our findings add to the evidence suggesting that face aftereffects are shaped by similar underlying processes to lower-level visual aftereffects – either a similar set of neural processes underlying adaptation, or a shared use of broader networks in the visual cortex (Leopold et al., 2005; Rhodes et al., 2007).

Finding the classic timecourse for expression aftereffects suggests that the effect is perceptual in nature, rather than being a result of decisional bias. An expression ‘aftereffect’ might potentially be explained by participants implementing a conscious or unconscious strategy to respond with the expression opposite the adaptor. It has been claimed that this sort of decisional bias might also follow a logarithmic timecourse (Storrs, 2015), but there is no current evidence that a decision bias would be affected by stimulus duration in this way, and it is not clear how this might occur. It seems more likely that the logarithmic timecourse of the expression aftereffect reflects the dynamics of neural processes that create changes in perception, as is the case for lower-level adaptation.

Anecdotal evidence also supports the perceptual nature of the effect. Following pilot testing and informal demonstrations of the effect, several of our colleagues have reported experiencing the test stimulus visibly changing as the aftereffect decays. During the debriefings for the present study, several participants expressed surprise when told that the test face was always the same image, and described seeing the test faces as varying in expression. These phenomena suggest that the measured aftereffect reflects a change in the
perceptual experience of facial expressions following adaptation, rather than a post-perceptual change in response bias.

In Experiment 1 we found significant expression aftereffects after just one second of adaptation. To our knowledge this is the shortest documented adaptation duration to successfully produce an expression aftereffect. We also found that the expression aftereffect was still present after 3200 ms of continuous exposure to the test face, at least for the longer adaptation durations (8s and 16s, possibly 4s). In Experiment 2 we found that expression aftereffects can persist over long gaps between adaptation and test (at least 32 seconds in duration), indicating that these aftereffects could potentially be retained over breaks in a social interaction. It is even possible that an expression aftereffect produced during one interaction could affect subsequent interactions with other individuals, especially given that expression aftereffects show (partial) transfer across identity (Skinner & Benton, 2012).

Given the long duration of the aftereffects found in Experiment 2, the expression aftereffect could also potentially persist over several changes of expression. However, we cannot be sure exactly how long the aftereffect would last when other faces or expressions appear after adaptation. In Experiment 2, we used a stimulus-free gap to vary the duration between adaptation and test. Kiani et al. (2014) found that facial identity aftereffects decay faster when other faces are seen between the adaptor and the test face. If the same is true for expression, the aftereffect may not last for as long as we have observed here if faces are present between adaptation and test. Given that there is both an identity-specific and identity-independent component of the expression aftereffect (Skinner & Benton, 2012), another interesting question is how the duration of the aftereffect is affected by intervening faces of the same identity as the adaptor (as in a one-on-
one conversation) as compared to intervening faces of a different identity to the adaptor (as in a conversation with multiple individuals).

Liberman, Fischer and Whitney (2014) have recently reported another effect of recent perceptual experience on the perception of faces. They found a serial dependence effect for facial identity, such that the perceived identity of a face is pulled towards the identity of faces seen in the several seconds previous. This effect operates in the opposite direction to the repulsive adaptation aftereffects reported here. Interestingly, this attractive effect of facial identity could be induced by viewing previous faces for as long as 1 sec. Here, we found that 1 sec of viewing a facial expression is enough to produce a repulsive aftereffect. Liberman et al. suggest that serial dependence helps to maintain visual stability that contributes to our experience of the continuity of objects. Given the dynamic nature of facial expressions compared to the more stable facial identity, it would be interesting to determine whether serial dependence operates at all for facial expressions, and if so, whether the ‘continuity field’, i.e., the length of time over which expressions have an attractive effect on the perception of subsequent expressions, is shorter for facial expressions than for facial identity.

We measured expression aftereffects here using a rating of perceived expression intensity, following Leopold et al. (2005) and Rhodes et al. (2007). In contrast, previous expression aftereffect studies have generally used a forced-choice expression-labeling task, and measured the aftereffect as a change in response thresholds or increase in proportion of responses opposite the adaptor (e.g. Burton et al., 2013; Fox & Barton, 2007; Hsu & Young, 2004; Pell & Richards, 2011; Skinner & Benton, 2010). The benefit of the present rating method is that participants are making a judgment about the intensity of a single expression, rather than having to make a choice between several expressions,
some of which may be readily confused with one another (e.g. fear and surprise, Gao & Maurer, 2010). Moreover, as expression intensity increases from very weak intensities, participants are able to identify that an expression is present before they are able to accurately identify what that expression is (X. Gao, personal communication, July 4, 2015). The forced choice method may therefore underestimate weak aftereffects, as participants may not be able to correctly identify what is nevertheless a change in their perception of the test face. The rating method may therefore be more useful in designs such as this one where very brief adaptation durations lead to weak aftereffects.

In conclusion, we have demonstrated that expression aftereffects follow the classic timecourse of build-up and decay seen in other face aftereffects and in lower-level vision. This finding provides further evidence of similarities in adaptive perceptual mechanisms between higher- and lower-level vision and reinforces the perceptual nature of expression aftereffects. We were able to produce expression aftereffects after only one second of adaptation, and longer adaptation periods produced aftereffects that were still present after more than three seconds of exposure to the test stimulus, and that remained after a gap between adaptation and test faces of more than 30 seconds. These findings suggest that expression aftereffects have the potential to affect our day-to-day social experiences.
Acknowledgements

This research was supported by the Australian Research Council (ARC) Centre of Excellence in Cognition and its Disorders (CE110001021), an ARC Professorial Fellowship to Rhodes (DP0877379), and an ARC Discovery Outstanding Researcher Award to Rhodes (DP130102300). We thank Andrew Skinner and Chris Benton for providing the stimuli. Ethical approval was granted by the Human Research Ethics Committee of the University of Western Australia.
References


Before calculating the aftereffect, we examined the baseline ratings from the start and end of each session to determine which would be the most appropriate baseline to use. If there were no difference between the two baseline blocks, it would be appropriate to take an average of ratings across the two blocks. If baseline ratings of expression intensity had increased significantly over the testing period, it would be most conservative to use only the post-adaptation baseline, so as to avoid over-estimating the size of the aftereffects.

We compared pre-adapt and post-adapt baseline ratings using a within-subjects ANOVA, with time (pre-adapt or post-adapt) session (first or second) and expression as within-subjects factors. We found no significant main effect of pre/post baseline block, $F(1, 22) = .018, p = .895, \eta^2_p = .001$, or session, $F(1, 22) = 1.67, p = .210, \eta^2_p = .070$. Not surprisingly, there was a significant main effect of expression, $F(1, 22) = 18.70, p < .001, \eta^2_p = .459$, indicating that the test face more closely resembled some expressions than others in the absence of adaptation. There was no significant session/pre-post interaction, $F(1, 22) = 2.57, p = .123, \eta^2_p = .105$, or session/expression interaction, $F(5, 110) = 1.38, p = .238, \eta^2_p = .059$. There was a significant pre-post/expression interaction, $F(5, 110) = 2.80, p = .020, \eta^2_p = .113$. There was no significant three-way interaction, $F(5, 110) = 0.94, p = .457, \eta^2_p = .041$.

The significant pre-post/expression interaction suggests that there might be significant differences in baseline ratings before and after the adaptation task for at least some expressions (Figure S1). We tested the significance of these differences using paired-samples t-tests. There was no significant difference between pre-adapt and post-adapt ratings for any of these expressions, all $ps > .008$ (bonferroni-corrected alpha). We therefore took an average of the ratings
from the two baseline blocks for each subject to give the final baseline rating for each expression.

*Figure S1.* Expression intensity ratings of the test (average) face at baseline, for each of the six target expressions, taken before adaptation task (Pre) and after adaptation task (Post). Ratings were made on a 7-point scale.
Chapter Five

General Discussion
General Discussion

This thesis aimed to investigate the visual representation of facial expression, and to uncover underlying perceptual processes that might contribute to the impressive expression perception ability displayed by most adults. To this end, the three sets of studies reported here used adaptation aftereffects to examine the way that facial expressions are visually represented. Our findings revealed that expression is represented in an opponent-coding system, relative to an implicitly-represented norm. We were unable to distinguish between two expressions, the average expression and the neutral expression, that have been suggested in the literature to occupy this position of expression norm. We found that expression aftereffects are perceptual in origin, and surprisingly persistent given the dynamic nature of expressions. We also found potential evidence for a functional benefit of this adaptability for expression perception.

Summary of the main findings

Opponent coding of facial expression

In Chapter 2, we found that an opponent coding model better explains the coding of facial expression than a model with an additional central coding channel. We used an adaptation paradigm that is able to distinguish between the two potential models of expression coding. Participants judged the expressions of faces on a fear/anti-fear trajectory in two adaptation conditions: adaptation to alternating fear and anti-fear, and adaptation to the center of the trajectory. If there was a central coding channel, then we would expect that adaptation in the alternating condition would widen the range of test levels that resembled the central expression, while adaptation in the central condition would narrow the range of test levels that resembled the central expression. Instead, we found that adaptation in both conditions resulted in a significant narrowing of the range of
test levels judged to resemble the central expression compared to baseline. This result supports a pure opponent-coded system for expression, without the presence of a central coding channel.

Previous research in support of an opponent coding for facial expression (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Skinner & Benton, 2010, 2012a) was unable to rule out the existence of an additional channel tuned to the center of the coded dimension, because both models produce the same predictions in the experimental designs used in these studies (Calder, Jenkins, Cassel, & Clifford, 2008; Lawson, Clifford, & Calder, 2009, 2011). Our study is the first to explicitly test whether a central coding channel is necessary to explain the coding of facial expression. Our support of an opponent coding model indicates that for facial expression the norm is represented implicitly, rather than explicitly. This implicit representation of the norm may contribute to the efficiency of the system, because neural resources are not expended in coding what is common to most expressions.

Our support for an opponent coding model of facial expression is based on predictions set out in previous studies that have used this paradigm (Calder et al., 2008; Lawson et al., 2009, 2011). The reasoning that leads to these predictions relies on the assumption that adaptation simply suppresses the responsiveness of the neural channels that are activated by the adaptor. However, this is not necessarily the only effect that adaptation might have on those neural channels. Adaptation might change the tuning function in other ways – for instance, by altering the slope of the tuning function. Predicting what might happen to the range of responses viewed as central following adaptation under these conditions becomes much more complex. Future research may be able to determine whether there is a set of assumptions about the effect of adaptation on the tuning functions
of the neural channels that would allow a central channel model to explain the pattern of results that we found here.

*Potential functional benefit for expression adaptation*

In Chapter 2, we found evidence for a potential functional benefit for adaptation in expression perception. Adaptation in both conditions narrowed the range of test levels judged to resemble the central expression (i.e. increased the range of expressions on either side of the center of the trajectory that were judged to resemble either fear or anti-fear) relative to baseline. This may be due to adaptation calibrating perception to the current range of stimulus variation. The greatest narrowing effect was seen after adaptation to the central expression. This central expression was an average expression that represents what is common to all expressions on the test trajectory, a feature that may have optimized the calibrating function of adaptation in this condition. As well as affecting perception of ambiguous expressions on the trajectory, we also found that adaptation to the central expression increased participants’ ability to identify strong fear and anti-fear expressions. The only other study to our knowledge that indicates a functional benefit for adaptation in expression perception is by Rhodes et al. (2015), who found a relationship between individual differences in expression adaptability and performance in an expression recognition task.

*No evidence that either the average or neutral expression is a better approximation of the expression norm*

In Chapter 3, we compared two expressions that have been put forward in the literature as potential norms for expression coding, the average expression and the neutral expression. We used an adaptation paradigm to determine which of these two expressions is a better approximation of the true norm of expression space. In an opponent-coded system, we expect larger aftereffects when adapting
along trajectories that pass through the norm than on trajectories that do not pass through the norm when trajectory length is equated (Rhodes & Jeffery, 2006). We measured adaptation aftereffects on trajectories made relative to the average expression and trajectories made relative to the neutral expression. We found significant aftereffects on all trajectories (except for the two trajectories on which we had difficulty estimating the response thresholds used to measure aftereffects, see Chapter 3). However, we did not find a significant difference in aftereffect size between the average and neutral trajectories. We were therefore unable to distinguish between the two potential norms. We suggest that this may be because the two expressions are relatively similar, and that both may be reasonable approximations of the norm. Both ‘norms’ are likely to be useful approximations of the true norm for the purposes of expression aftereffect research, because they both produce anti-expressions that are capable of generating expression aftereffects.

It is important not to over-interpret the results of this study, which showed no significant difference between the two potential norms. Our sample size, while fairly typical for face aftereffect research, was relatively small. As a result, our study may have been underpowered to find anything other than a large effect. It is difficult to determine what effect size we might have expected to find in this study, however, because the difference in aftereffect size between the two conditions depends on the relative proximity of the two approximate norms to the true norm of the space. If one expression is only slightly closer to the norm than the other, then the effect size will be very small, and the difference therefore difficult to detect. Such a small difference, however, is unlikely to be of any theoretical importance.
Expression aftereffects show the classic timecourse pattern found for perceptual aftereffects

In Chapter 4, we showed that the expression aftereffect follows the same classic timecourse pattern of logarithmic build-up and exponential decay found in low-level perceptual aftereffects (Krauskopf, 1954; Sekuler, 1975; Wolfe, 1984). We varied the duration of adaptation and test exposure in an expression aftereffect task, and measured the aftereffect as the strength of the participant’s impression of the expression opposite the adaptor (measured on a seven-point rating scale). We found that aftereffects increased logarithmically with adaptation duration and decreased exponentially with test duration, following the timecourse pattern found for low-level perceptual aftereffects. This pattern has also been demonstrated for identity and figural face aftereffects (Leopold, Rhodes, Müller, & Jeffery, 2005; Rhodes, Jeffery, Clifford, & Leopold, 2007). It is hard to see how this timecourse could be caused by non-perceptual explanations for the expression aftereffect, such as response biases created by exposure to the adaptor. We therefore argue that this classic timecourse pattern is evidence for a perceptual locus for expression aftereffects. Participants moved their eyes freely during adaptation and there was a size change between adapt and test, so the timecourse pattern also cannot be explained by retinotopic low level adaptation. This finding is important, because the use of expression aftereffects to investigate the visual representation of facial expression relies on the assumption that these aftereffects reflect adaptation of perceptual mechanisms.

Expression aftereffects are surprisingly persistent

In Chapter 3 we also explored the persistence of expression aftereffects, and found that these aftereffects are relatively long-lasting: just eight seconds of adaptation produced an aftereffect that was still present after an adapt-test delay of
32 seconds (the longest delay that we tested). The adaptation durations tested here are similar to the duration of the spontaneous expressions that have been observed in other studies (Frank, Ekman, & Friesen, 1993; Pfister, Li, Zhao, & Pietikainen, 2011). Given the persistence of the expression aftereffects reported here, it is therefore likely that adaptation has a pervasive effect on the day-to-day experience of facial expressions.

**Integrating the main findings**

The studies summarized above were linked by the over-arching aim of this thesis: to better understand the mechanisms that underlie the visual representation of facial expressions. Below we outline the major themes that emerge from these findings.

*Norm-based coding of expression*

Our findings here build on previous evidence suggesting that expression is coded relative to a norm, by showing how that norm may be represented. We found that expression is likely to be coded in a two-pool opponent-coded system rather than a central-channel coding system (Chapter 2). In an opponent coding system, the norm is represented implicitly as the point at which both pools of neurons respond equally. This implicit coding of the norm is efficient, because neural resources are reserved for coding what is distinctive. Our evidence against a central-channel model of expression coding contrasts with previous evidence that a central channel is used in the representation of gaze direction and head orientation (Calder et al., 2008; Lawson et al., 2011). These two types of face-related information are both dynamic, like expression. However, they are unlike expression in that they are both direction-related. For both gaze direction and head orientation the state represented by the center of the dimension (face pointed at the viewer, direct gaze) is important and meaningful. In contrast, for expression
the middle of the dimension is likely to be a less-important ambiguous expression, with meaningful information coming from expressions that deviate from the central state. For these reasons, it is perhaps unsurprising that expression is represented in a different way to gaze direction and head orientation.

We then attempted to determine whether an average or neutral expression is a better approximation of this expression norm (Chapter 3). We were unable to distinguish between an average and a neutral expression in this experiment. There are reasons to believe that the expression norm in an opponent coding system should be an average expression, because an average norm allows the system to avoid coding the aspects of an expression that are common, focusing coding on the most diagnostic aspects. An average expression norm could also be updated over time, so that it represents the central tendency of the current range of expressions around which distinctions need to be made. Finally, it is easier to see how an average norm could be maintained in an opponent-coded system. The norm lies at the center of the dimension, and to maintain its position expressions must be experienced that lie on either side of this norm. In the case of a neutral norm, some of these expressions may not be possible, because the neutral expression lies at one extreme of all possible muscle movements (i.e. it is the face when all muscles are relaxed: Ekman, Friesen, & Hager, 1978). For instance, it is difficult to see how an expression could be experienced that perfectly opposes an expression of surprise relative to neutral, because in order to balance the open mouth of surprise the opposing expression would need to have a mouth that is “more closed” than the already-closed mouth of the neutral expression. This problem explains the visual artifacts around the mouth that led us to exclude the neutral-referenced anti-happiness and anti-surprise expressions from the experiment described in Chapter 3. Although these artifacts are most visible
around the mouth, this is not the only place where creating opposing expressions is problematic. For instance, what expression might oppose the raised cheeks of a smiling expression relative to neutral, where the cheeks are already completely relaxed?

Given the arguments presented above for an average expression being a superior contender for the position of expression norm, it is perhaps surprising that we were unable to reveal a difference between the two expressions with our aftereffect paradigm. However, it may be the case that the average expression that we tested here (which is only one example of an “average” expression – see Chapter 3 for a further discussion) and the neutral expression are both reasonably close in expressions space to the true norm. The two expressions are somewhat similar in appearance, and both appear relatively unexpressive (although see Bimler & Kirkland, 2001; Gao, Maurer, & Nishimura, 2010; Lee, Kang, Park, Kim, & An, 2008; J. A. Russell & Bullock, 1985; J.A. Russell & Bullock, 1986, which indicate that the neutral expression can be perceived as somewhat negative). Although we have argued above that neutral is not at the center of any possible muscle movements, a neutral expression may lie nearer the middle of more broadly-defined deformations of the face: for instance, the raised brow of surprise and the lowered brow of anger. To the extent that the muscles of the faces are arranged in opposing pairs, the neutral expression will still lie between the expressions created by these paired muscles. This would explain why neutral does bear some resemblance to the average expression, and suggests that neutral could still confer some efficiency as a norm against which expressions are represented.

What do we mean by “expression”?

When we ask what might constitute the norm for expression coding, it is important to consider what we mean by the term “expression”. Ekman identified
six basic expressions of emotion (anger, disgust, fear, happiness, sadness and surprise) that are easily and reliably recognized across cultures (Ekman, 1970). Our stimuli throughout this thesis are derived from these six expressions of emotion. However, we would like to make it clear that we do not believe that these are necessarily the only aspects of face perception that are processed by the expression mechanisms that we explore throughout this thesis. There are a great number of other “expressions” (or facial postures) that can be produced through dynamic movements of the facial muscles (Cook, Matei, & Johnston, 2011). These expressions are recognizable, interpretable and useful in social interactions. For instance, there are expressions that reflect internal states other than these six basic emotions, such as thoughtfulness, confusion and boredom. Other expressions may be intentionally displayed as communicative signals, such as a wink or raised eyebrow. Finally there are other facial positions that are less obviously part of the set of facial expressions, such as the movements of the lips with speech, which are also recognizable and provide social information. It could well be most efficient for all of these types of expression to be represented together at the visual level. If our representation of facial expressions includes expressions beyond the basic expressions of emotion, then the position of the norm should also reflect these other sources of expression variation.

*Expression perception is adaptable*

Throughout the studies presented here, we confirm that the perception of facial expressions is adaptable, such that the appearance of an expression depends on what has been previously experienced. In Chapter 4, we demonstrate that this expression aftereffect is likely to be perceptual in origin. This finding is particularly important because other expression adaptation studies (including those reported in this thesis) rely on the assumption that expression aftereffects
tap perceptual processes, and are therefore able to inform us about the mechanisms of that perception. More generally, the classic timecourse pattern for expression aftereffects suggests that the visual coding of facial expressions is flexible, a feature of the coding system that may have functional benefits. Indeed, we found some support for this proposition. In Chapter 2, adaptation to the central expression on a fear/anti-fear trajectory improved participant’s ability to identify both fear and anti-fear expressions. This might occur because the sensitive range of the coding system is shifted with adaptation, improving sensitivity around the adapted set of expressions. In Chapter 4, we also found that expression aftereffects are surprisingly persistent. A long-lived expression aftereffect could help the visual system to integrate information about the current range of expressions that are seen, so that the adapted state reflects not just the most recently seen expression, but rather the range of expressions across which discrimination is currently necessary (Kloth & Rhodes, In Press).

*Distinguishing between perceptual and post-perceptual processes*

Our finding in Chapter 2 that adaptation to the central expression improved identification of the fear and anti-fear expressions may indicate a functional benefit of adaptation, as discussed above. However, this change in responses does not necessarily reflect increased sensitivity. It is also possible that viewing the adaptors altered participants’ decision criteria, perhaps by offering a reference point against which to judge the test faces. Distinguishing between perceptual and post-perceptual processes is a challenging problem throughout perception research (Morgan, Dillenburger, Raphael, & Solomon, 2012), and in the case of these data the two cannot be separated. However, there are alternative psychophysical methods that will allow us to more clearly see whether adaptation has an effect on sensitivity, separate from confounds like response bias. One such
method is a three-alternative odd-one-out task (e.g. Dakin & Omigie, 2009). This type of task does not require that participants categorize the stimuli in any way – they must simply identify the stimulus that differs from the others. The task therefore measures only whether or not a participant can detect the differences between the stimuli. We are currently conducting research that uses this odd-one-out method to describe how expression sensitivity varies with distance from the mean. We aim to discover whether expression sensitivity is greatest around the average expression, as has been demonstrated for identity (Dakin & Omigie, 2009). We hope that we will then be able to examine whether sensitivity to expressions is improved by adaptation, as is suggested by our findings in Chapter 2.

The relationship between identity and expression perception

Overall, we find evidence that the visual representation of facial expressions uses similar mechanisms to the visual representation of facial identity. One possibility to consider is that expression and identity may in fact be represented in a single, shared system. In two influential models of face perception, identity and expression are separated early in visual processing (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000). This theorized separation was based on a number of neuropsychological findings, including a double dissociation between expression and identity recognition in patients with lesions or other disruptions to face processing. However, it has more recently been suggested that this dissociation could occur later in processing, and that expression and identity information may be processed together at the level of visual representation (Calder, 2011; Calder & Young, 2005; Rhodes et al., 2015).

Regardless of whether or not expression and identity are visually represented together, disentangling these two types of information remains a
complex task. Determining that two images show similar but different identities, as opposed to one identity with different expressions, involves determining which dynamic changes to a face are possible or likely, and which differences in face images are unlikely to be created by typical dynamic changes to a face of a single identity. Calder, Burton, Miller, Young, and Akamatsu (2001) used a principal components analysis to describe the variation between face images, and showed that there are separable components for distinguishing between identities and distinguishing between expressions. The dimensions on which faces are coded are likely based on the learned statistical properties of face images. If this is the case then there may be coding dimensions that describe the variation that can be achieved through changes in expression, independent of identity, that allow us to separate the two types of information.

At the same time, there appears to be a degree of overlap to expression and identity information in a face. Rhodes et al. (2015) found a positive correlation between expression and identity aftereffects, and found that this shared adaptability significantly predicted performance on both expression and identity recognition tasks. This common variation is evidence for at least some coding dimensions that are relevant to both expression and identity perception. Calder et al. (2001) also found some overlap between the identity-relevant and expression-relevant components extracted by their analysis. This overlap may explain why when viewing a face (particularly in a still photograph) we will occasionally confuse what is structural information and what is dynamic. For instance, some faces have structural qualities that resemble expressions (a face that resembles anger even at rest because of a low, drawn-together brow, etc.). Viewers will tend to make trait judgments about those faces that are influenced by the apparent expression: for instance, a person whose face subtly resembles anger will tend to
be judged as threatening (Said, Sebe, & Todorov, 2009). Additionally, Andrews, Jenkins, Cursiter, and Burton (2015) have demonstrated that participants have difficulty determining how many different identities are present in images of similar-looking unfamiliar individuals. These images are naturalistic, and vary in characteristics such as viewpoint, lighting, expression, makeup, and so on. Participants will tend to overestimate the number of identities present, suggesting that they confuse dynamic changes (including expression as well as viewpoint, lighting, etc.) with structural differences. In contrast, participants can perform this task accurately if the individuals shown are familiar. This advantage for familiar faces suggests that as well as a general representation of the transformations that are produced by dynamic changes to faces, we may also store information about the range of variation that can be achieved by a particular individual’s face, improving our ability to separate dynamic and structural information for familiar individuals.

**Advantages to our statistical methods: The mixed logit model**

In Chapters 2 and 3 we use a mixed logit model, a type of generalized linear mixed model, to analyse our data. This type of analysis is common in fields such as economics, but is under-utilised in psychophysics research, despite being ideally suited to analyzing psychophysical data. The design of many psychophysical experiments requires that we plot participants’ response functions across various levels of a test stimulus (in our case, levels of expression strength), and then test for changes in that response function. Individual participants may have different response functions, so it is not ideal to fit functions to the group data. The traditional way of dealing with this type of data is to fit functions for each participant individually, estimate the values of interest for each participant (e.g. a 75% response threshold), and then enter those estimated values in a
secondary analysis such as an ANOVA. A limitation of this approach is that it
treats the estimated values as if they were true measurements, discarding
information about the quality of the function fits (Moscatelli, Mezzetti, &
Lacquaniti, 2012).

A generalized linear mixed model is ideal for this kind of data, because a
single model can be fit simultaneously to all participants’ data. Rather than fitting
a single fixed value for each parameter describing the function, we are able to
allow some parameters to vary between participants. The standard errors of any
values estimated from this model (e.g. thresholds) take into account the quality of
the fit of the model to each participant’s data. These standard errors can then be
used to calculate the significance of experimental effects. This method of analysis
gives greater statistical power than the traditional two-stage approach (Moscatelli
et al., 2012).

**Future directions**

Our findings suggest several avenues for future study. One area that
should be explored further is the persistence of the expression aftereffect. In
Chapter 4, we found that the expression aftereffect was still present after a 32
second delay between adaptation and test. We suggest that this persistence may
allow adaptive perceptual processes to integrate information over time, giving a
norm that better represents the current range of expression variation. However, it
should be noted that during the delay between adaptation and test, participants
were not exposed to any face stimuli. We should therefore consider whether at
least a part of the persistence reported here might be due to storage effects. In the
case of some low-level aftereffects, like the motion aftereffect, introducing a
stimulus-free period immediately after adaptation makes it possible to measure a
significant aftereffect long after it would otherwise have decayed (see van de
Grind, van de Smagt, & Verstraten, 2004 for a brief review). Storage is classically demonstrated using a period of complete darkness following adaptation (MacKay & MacKay, 1977), but recently it has been shown that for some aftereffects, it is sufficient that there are no stimuli present after adaptation that have the same properties as the adaptor (Thompson & Wright, 1994). Our face-free delay period may therefore have allowed storage of the expression aftereffect. It is possible that the aftereffect might dissipate sooner if other faces are present immediately after adaptation. We may therefore have overestimated the usefulness of the expression aftereffect for integrating expression information over time.

To clarify whether the aftereffect duration we measured in Chapter 4 was influenced by storage effects, it would be useful to repeat the persistence experiment, but with intervening faces included in the delay between adaptation and test. We would then be able to determine whether the presence of face stimuli after adaptation causes the aftereffect to decay at a faster rate, and if so, how long the aftereffect persists with exposure to faces, as would typically be encountered in a social interaction. A similar method has recently been used to show that the identity aftereffect decays faster when faces are present following adaptation (Kiani, Davies-Thompson, & Barton, 2014). This hastened decay might be explained by the system re-adapting to the intervening face stimuli (van de Grind et al., 2004). It would therefore also be interesting to determine whether the speed with which the expression aftereffect decays depends on the expression of the faces seen after adaptation. If the decay is due to the system re-adapting, then we might expect that aftereffects will decay more quickly if faces seen after adaptation show a strong version of the expression opposite the adaptor, as opposed to a weaker expression. Another variable of interest is the identity of the intervening faces. We speculate that intervening faces that match the identity of
the adaptor might cause a faster decay than different-identity adaptors. This arrangement would allow an adapted state developed during an interaction with one individual to be maintained at least to some extent over brief interruptions from other individuals (e.g. glances at a third person during a conversation).

Another area in which progress might be made from our current findings is to increase the naturalism of our stimuli. The stimuli used here were chosen because they are easily controlled, and have been established to produce reliable expression aftereffects in previous literature (Burton et al., 2013; Skinner & Benton, 2010, 2012a, 2012b). However, they are very standardized, somewhat unnatural images. For instance, all of the faces show the same identity, made from a morph of fifty individuals. This set of individuals included both males and females, resulting in an androgynous morph, another feature that is uncommon in the expressions we typically experience. The images were also shown in greyscale. This colour choice helps to conceal any strangeness in colour created by the morphing process, but it does not correspond to the way that we typically see faces. More generally, the faces that we see in everyday life show much more variation in viewing angle, lighting, makeup, facial hair, etc. Finally, the face images were static, and unless we are looking at a photograph, expressions are generally seen in motion. As much as possible, it would be good to establish that our findings here generalize to more naturalistic stimuli. For some of the characteristics described above, this would not be difficult – for instance, Skinner and Benton (Skinner & Benton, 2012a) have demonstrated expression adaptation across different identities, and describe a method for controlling the expression on the face while letting identity vary.

Introducing moving stimuli is a more complicated problem, because it is not possible to produce moving anti-expression adaptors. One partial solution may
be to adapt participants to a typical expression in motion (e.g. an unfolding smile), and then to use static anti-expressions as our test stimuli (a method used for facial identity adaptation by Jeffery, Petrovski, & Rhodes, 2014). The difficulty here would be in training participants to recognize and respond to the unfamiliar anti-expressions. However, this is not unachievable – participants were able to learn and identify the anti-fear expression in Chapter 2. Besides allowing the use of moving adaptors, adapting participants to expressions rather than anti-expressions may also have the benefit of maximizing the aftereffects produced by adaptation.

In Chapter 2 we observed that adapting to alternating fear and anti-fear faces caused an overall shift in the response functions towards fear, suggesting that the system adapted more strongly to this familiar expression than to unfamiliar anti-fear. Using adaptors that produce maximal aftereffects could improve our ability to detect subtle differences between conditions in future experiments.

Conclusions

The research described in this thesis aimed to contribute towards a better understanding of the way that facial expressions are visually represented. Adaptation methods were used to clarify what mechanisms might code facial expression, using findings from identity representation research to guide our investigation. We confirmed that the visual representation of facial expressions is adaptable, and supported the widespread assumption that expression aftereffects are perceptual in origin. We found evidence that might suggest a functional benefit of this adaptable coding, a finding that had only previously been suggested by a single study. We also found that expression is coded in an efficient two-pool opponent system, ruling out for the first time a central-channel model. These findings mirror what has been found for facial identity, and are consistent with a common representation for these two types of information. It is hoped that the
ideas presented in this thesis shed light on the processes that contribute to our impressive ability to read the expressions of others, an important social skill.
References:


Appendix One

An Initial Attempt to Find the Center of Expression Space
Appendix

An initial attempt to find the center of expression space

Note: This experiment was originally designed to take the place of the experiment reported in Chapter 3. However, there were problems with the design (discussed in detail below) that made this experiment unsuitable for answering our research question in that chapter. The motivation and most of the methods are the same between the two experiments, and as a result the majority of the introduction and stimulus development sections here are the same as those found in Chapter 3. While there were design problems with this experiment, there are still some interesting observations that we can make about these data. It may also be useful for researchers who (like the authors) are not experienced psychophysicists to consider the problems with the logic of this design, as these issues were not immediately apparent to us before analyzing the data.
Facial expressions are an important source of social information, providing cues to the emotions and intentions of others. Despite the similarity of faces as visual patterns, most people are able to quickly and easily interpret expression from a single glance at a face. Norm-based coding is one mechanism that may contribute to our proficiency in making judgments about facial expressions (Rhodes & Leopold, 2011). In this type of visual coding, faces are represented in a multi-dimensional space in which each dimension represents a way in which faces are perceived to vary. The ‘norm’ is located at the center of the space, and acts as a reference point against which the positions of other faces are coded. This type of coding is efficient, because it focuses coding on what is diagnostic about a face rather than what is common to most faces, and flexible, because the system is calibrated by experience (Rhodes & Leopold, 2011; Webster & MacLeod, 2011).

Adaptation aftereffects are a useful tool for investigating how facial expressions are visually represented. These aftereffects occur when exposure to one stimulus reduces the responsiveness of the neurons that code it, altering the neural response elicited by subsequent stimuli (Webster & MacLeod, 2011). For instance, adapting to constant motion in one direction will temporarily make a stationary object viewed immediately afterwards appear to move in the other direction (Anstis, Verstraten, & Mather, 1998). Adaptation can be induced by viewing faces: for example, adapting to faces that have been distorted so that the features are contracted into the center of the face will make an undistorted face appear expanded (Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Webster & MacLin, 1999), and adapting to male faces makes an androgynous face appear more female (Webster, Kaping, Mizokami, & Duhamel, 2004). These face aftereffects are likely to reflect adaptation of higher-level face
representations, not just low-level, retinotopic adaptation, because they persist over changes in the size, colour, contrast and orientation of a face, and when free eye movements are allowed (Rhodes et al., 2005; Watson & Clifford, 2003; Yamashita, Hardy, De Valois, & Webster, 2005; Zhao & Chubb, 2001).

Similarly, adaptation to a particular facial expression will bias perception of subsequent expressions towards the opposite expression (Burton, Jeffery, Skinner, Benton, & Rhodes, 2013; Cook, Matei, & Johnston, 2011; Juricevic & Webster, 2012; Skinner & Benton, 2010, 2012). The opposite of an expression is its anti-expression, which has attributes that differ from the norm to the same extent as the original expression, but in the opposite direction. For instance, where fear has raised brows relative to the norm, anti-fear has lowered brows (Figure 1). In the same way that adapting to a contracted face makes subsequent faces appear expanded, adapting to an anti-expression makes subsequent faces look more like the original expression (Burton et al., 2013; Cook et al., 2011; Juricevic & Webster, 2012; Skinner & Benton, 2010, 2012). These effects persist over changes in size and when free eye movements are allowed (Burton et al., 2013; Cook et al., 2011; Fox & Barton, 2007; Hsu & Young, 2004; Skinner & Benton, 2010, 2012), again suggesting that they tap higher-level face processing.
Evidence from several aftereffect studies suggests that facial expressions are coded relative to a norm (Burton et al., 2013; Cook et al., 2011; Skinner & Benton, 2010, 2012). Norm-based coding can be implemented by a two-pool opponent coding system. In this type of coding, two pools of neurons code each dimension of perceived expression variation, one responding maximally to each end of the dimension. The norm is represented by the point where both pools respond equally. Converging evidence from a range of paradigms supports a two-pool coding system that represents facial expressions relative to a norm (Burton, Jeffery, Calder, & Rhodes, 2015; Burton et al., 2013; Cook et al., 2011; Skinner & Benton, 2010, 2012).

Although this evidence suggests that a norm expression plays an important role in the representation of facial expressions, there is currently no consensus on precisely what this norm expression might be. Studies investigating norm-based coding of facial expressions have used one of two different potential expression norms: an average expression (the central tendency of facial expressions), or a neutral expression (when the face is at rest) (Figure 2). The average expression is
made by taking an average from a set of expression images, of either the basic emotion expressions (Burton et al., 2013; Skinner & Benton, 2010, 2012) or more varied facial postures (Cook et al., 2011). This average expression is intended to lie near the mean on each of the various dimensions on which expressions are coded. An average expression is analogous to the type of norm often used in facial identity research (where the concept of face space was originally developed), which is generally created by averaging together a large set of different identities (for a review see Rhodes & Leopold, 2011). For expression coding, the benefit of an average as the norm is that the average represents what is common to most expressions, maximizing the efficiency granted by norm-based coding.

Other researchers have used a neutral expression as the expression norm (Juricevic & Webster, 2012; Rutherford, Chattha, & Krysko, 2008; Rutherford, Troubridge, & Walsh, 2011). Webster and MacLeod (2011) suggest that the neutral expression has the property of appearing unexpressive or “psychologically neutral,” analogous to the way that white, the norm of opponent-coded colour space, appears neutral. Indeed, it is commonly assumed that a neutral face is unexpressive, and so can be used as a baseline against which other expressions are compared. For example, tests of expression sensitivity create expressions of varying strengths by morphing between a given expression and neutral (e.g. Gao & Maurer, 2009, 2010). Likewise, the neutral expression has been used to gauge the baseline response in neural imaging studies looking at responses to facial expressions (Sprengelmeyer et al., 1998; Kesler-West et al., 2001; Pessoa et al., 2002; Kilts et al., 2003).

Not all evidence, however, suggests that the neutral face is unexpressive. In an implicit emotion evaluation task, Lee, Kang, Park, Kim, and An (2008) found that participants responded to neutral expression stimuli in the same way
they did to negative expression stimuli. Moreover, multidimensional scaling of facial expressions in several studies has placed neutral expressions away from the center of the space, nearer to other emotion expressions, such as sadness or sleepiness (Bimler & Kirkland, 2001; Gao, Maurer, & Nishimura, 2010; J. A. Russell & Bullock, 1985; J.A. Russell & Bullock, 1986). These findings suggest that the “neutral” expression might be better regarded as another expression, with negative rather than neutral valence. If so then this undermines the principle reason to believe that it should be the norm of expression space.

Our aim was to determine more directly which expression, average or neutral, is a better approximation of the norm of expression space. To do so, we exploited the finding that in a norm-based system, adapt-test pairs that lie opposite one another relative to the norm produce larger aftereffects than adapt-test pairs that are not opposite one another (Leopold, O'Toole, Vetter, & Blanz, 2001), even when perceptual distance between the pairs is equal (Rhodes & Jeffery, 2006). This effect occurs because adaptation selectively biases perception towards the opposite of the adaptor across the norm. This property has been used to identify the best approximation of the norm for different face categories in facial identity research. For example, adapt-test pairs that lie opposite one another relative to a race-specific average (Asian or Caucasian) produce larger identity aftereffects than adapt-test pairs opposite one another relative to a mixed-race average, indicating that the norms used to code identity are race-specific (Armann, Jeffery, Calder, Bülthoff, & Rhodes, 2011). Similarly, adapt-test pairs that are opposite on another relative to a sex-specific average produce larger identity aftereffects than adapt-test pairs opposite one another relative to an androgynous average, indicating that identity is coded relative to sex-specific norms (Rhodes et al., 2011). Following this logic, expression aftereffects should be larger on
trajectories that pass through the true expression norm (opposite trajectories) than on trajectories that do not (non-opposite trajectories). Therefore, if one of our two potential norms (neutral or average) is a better approximation of the true expression norm, then we will see larger aftereffects on trajectories that pass through that expression (with adapt-reference face pairs equated on perceptual similarity).

Here we adapted and tested participants on trajectories that pass either through the average expression, or through the neutral expression. We refer to these two potential approximations of the norm as ‘reference faces’ because we used them as reference points around which we created expression trajectories. If one of these two expressions is a good approximation of the true norm, we expected to see larger aftereffects on trajectories that pass through that expression (opposite trajectories) than on trajectories that pass through the other expression (non-opposite trajectories).

**Method**

**Participants.** Participants were 24 Caucasian adult volunteers (four males) from the UWA community, aged 17-25 years (M=19.6 years, SD=2.9 years). Participants either took part for course credit, or were reimbursed $10 for their time and travel expenses. One additional participant was tested but was excluded from analysis because he was unable to complete the training stage.

**Stimuli**

Stimuli were developed from those used by Skinner and Benton (2010). The original stimuli were constructed from photographs of 25 male and 25 female faces posing the six basic expressions (happy, sad, angry, fearful, surprised and disgusted) and a neutral expression. For each expression the 50 images were combined using Psychomorph morphing software (Tiddeman, Burt, & Perrett,
(2001) to create a single composite image of that expression. The composite of the neutral expressions created by this process was used as our neutral reference face. To produce an average reference face, the seven composite expressions were combined into a single composite image (Figure 2).

![Figure 2. The two 0% reference faces, neutral (left) and average (right).](image)

Each expression trajectory was created by morphing an expression through one of the reference faces and out beyond it to an anti-expression. The features of the anti-expression differed from the reference face to the same extent as those of the original expression, but in the opposite direction. Expressions on these morph trajectories can be identified using percentages, where 100% indicates the original expression, 0% indicates the reference face and -100% indicates the anti-expression.

Because the neutral expression has a closed mouth, anti-expressions created by morphing along a trajectory from an open-mouthed expression through neutral contained some visual artifacts around the mouth. We avoided using the expressions that showed the worst of these artifacts (happiness and surprise), and chose anger, fear, disgust and sadness as our four target expressions. There were
two trajectories for each expression (one made with the average reference face, one with the neutral reference face), giving eight trajectories in total.

To meaningfully compare the size of the aftereffects produced by the neutral-referenced and average-referenced adaptors, the perceptual distance from the adaptors to the reference faces should be equal for both sets of trajectories (Rhodes & Jeffery, 2006), because more extreme adaptors produce stronger aftereffects (Burton et al., 2013; Skinner & Benton, 2010, 2012). To equate these distances, a separate group of participants rated the perceived similarity of the anti-expressions to their respective reference faces (see Chapter 3, Supplementary Materials A). Participants viewed each anti-expression/reference face pair one at a time, and rated how similar the two faces looked on a seven-point scale. We used these ratings to select adaptor strengths for each trajectory such that for each expression the average-referenced anti-expression adaptor was judged to be as similar to the average reference face as the neutral-referenced anti-expression adaptor was to the neutral reference face. The final similarity-matched adaptors are shown in Figure 3.
<table>
<thead>
<tr>
<th>Expression Composites</th>
<th>Anger</th>
<th>Fear</th>
<th>Disgust</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

| Average-Ref.          | ![Image](image5.png) | ![Image](image6.png) | ![Image](image7.png) | ![Image](image8.png) |
|                       | -100% | -100%| -100%   | -100%   |

| Full Strength, Anti-Expressions | ![Image](image9.png) | ![Image](image10.png) | ![Image](image11.png) | ![Image](image12.png) |
|                                 | -100% | -100%| -100%   | -100%   |

| Neutral-Ref.             | ![Image](image13.png) | ![Image](image14.png) | ![Image](image15.png) | ![Image](image16.png) |
|                         | -100% | -100%| -100%   | -100%   |

<table>
<thead>
<tr>
<th>Average-Ref.</th>
<th><img src="image17.png" alt="Image" /></th>
<th><img src="image18.png" alt="Image" /></th>
<th><img src="image19.png" alt="Image" /></th>
<th><img src="image20.png" alt="Image" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptually Equated, Anti-Expressions</td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td>Neutral-Ref.</td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
<td><img src="image28.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td>-100%</td>
<td>-100%</td>
<td>75%</td>
<td>-100%</td>
</tr>
</tbody>
</table>

Figure 3. Expression stimuli. Top row are the full-strength expression composites used to generate the anti-expressions. Below are the full-strength anti-expressions made relative to the average and neutral anti-expressions. Bottom rows are the final set of eight perceptually equated anti-expression adaptors. These adaptors were selected so that for each expression the average-referenced and neutral-referenced adaptors were matched for rated similarity to their respective reference faces.
Test faces were the two 0% reference faces (Figure 2). We also included weaker versions of the target expressions, at 60% strength, as test expressions to keep participants motivated during the task. These were taken from both the neutral and the average trajectories for each expression. Adaptors subtended a viewing angle of $8.8^\circ$ by $11.9^\circ$ when viewed from 50 cm. Test expressions were displayed at 75% of the size of the adaptors (subtending a viewing angle of $6.6^\circ$ by $8.9^\circ$) and free eye movements were allowed to reduce the effect of low-level adaptation.

**Procedure.** In each trial of the adaptation task, participants first viewed an adaptor, and then saw and identified a test expression. Participants began the testing session by learning to identify the target expressions in a training task.

**Training.** First participants were shown the four target expressions, one at a time, at 100% strength and for an unlimited duration, and identified them using marked keyboard keys. Participants viewed and identified the expressions in a continuous sequence, with every run of the four expressions reordered randomly. Participants continued until they identified two consecutive runs with 100% accuracy (number of repetitions to reach criterion ranged from 1-5, $M = 1.58$, $SD = 0.98$). Participants then moved on to identify the neutral- and average-referenced 60% versions of each target expression with unlimited viewing duration. Participants again viewed and identified the expressions in a continuous sequence, with every run of the eight expressions reordered randomly. Participants continued until they identified two consecutive runs with 100% accuracy (number of repetitions to reach criterion ranged from 1-15, $M = 2.46$, $SD = 2.79$; two participants performed 11 repetitions and one participant performed 15 repetitions, but these participants did not perform unusually in the final
training sequence and removing their data did not affect the outcome of analyses, so they were retained in the sample).

Participants were then introduced to the shorter viewing time that would be used in the adaptation task. An expression was shown for 400msc, and was then replaced with a response screen. Participants practiced this procedure with the four 100% expressions. They then identified the eight 60% expressions with this shorter viewing duration. They were required to identify two consecutive runs of the eight expressions with at least 75% accuracy (number of repetitions to reach criterion ranged from 1-3, $M = 1.06$, $SD = 0.32$). Finally, participants practiced identifying an even weaker set of expressions (20% strength) to accustom them to performing the task with more ambiguous stimuli. There was no performance criterion for this final set of expressions.

**Adaptation Task.** In each trial, an adaptor was shown for 8000msc, followed by a 150ms blank ISI. The test expression was shown for 400msc. After another 150ms ISI, a response screen was shown, and participants identified the expression they had just seen. Participants pressed the space bar to initiate the next trial.

The testing session was split into two blocks, one testing the average-referenced trajectories and one testing the neutral-referenced trajectories. The order of these blocks was counterbalanced between participants. Each of the four anti-expression adaptors within a block appeared in 10 trials. In eight of those trials the test expression was the 0% reference face, and in two the test was a 60% expression. These 40 adapt trials were intermixed with 40 baseline trials, in which the adaptor was replaced with the grey oval. Like the adapt trials, 32 of the baseline trials used the 0% reference as the test expression and 8 used the 60% expressions. Trials were presented in random order within each block. In total
participants completed 160 trials (80 per block). Participants took breaks every 40 trials and testing took approximately 30 minutes.

Results

The aftereffect is measured as the change in participant’s tendency to perceive the 0% test face as the target expression following adaptation relative to baseline. We calculated aftereffects separately for each of the eight test trajectories for each participant. First we calculated each participant’s baseline response rates for each of the two 0% test faces (average and neutral). These were the proportion of no-adapt trials in which participants applied each target expression to the 0% test face (Table 2). We then calculated the response rate after adaptation for each of the eight trajectories. This was the percentage of trials in which participants responded that they saw the relevant target expression after adapting to that trajectory’s adaptor (e.g., the percentage of trials in which participants responded “angry” after adapting to neutral-referenced anti-angry). We then subtracted the relevant baseline response rate (e.g., baseline “anger” responses to 0% neutral for the neutral-referenced anger trajectory) from the adapted response rate to derive the aftereffect for that trajectory.
Table 2: Baseline response rates: the proportion of trials in which the average and neutral expressions were labeled as each of the target expressions with no adaptation.

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Neutral</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
<td>Post-Adapt</td>
<td>Baseline</td>
<td>Post-Adapt</td>
<td>Baseline</td>
<td>Post-Adapt</td>
<td>Baseline</td>
<td>Post-Adapt</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SE$</td>
<td>$M$</td>
<td>$SE$</td>
<td>$M$</td>
<td>$SE$</td>
<td>$M$</td>
<td>$SE$</td>
</tr>
<tr>
<td>Anger</td>
<td>.16</td>
<td>.03</td>
<td>.37</td>
<td>.06</td>
<td>.34</td>
<td>.05</td>
<td>.56</td>
<td>.08</td>
</tr>
<tr>
<td>Fear</td>
<td>.39</td>
<td>.05</td>
<td>.78</td>
<td>.06</td>
<td>.04</td>
<td>.01</td>
<td>.08</td>
<td>.03</td>
</tr>
<tr>
<td>Disgust</td>
<td>.26</td>
<td>.04</td>
<td>.52</td>
<td>.07</td>
<td>.03</td>
<td>.01</td>
<td>.06</td>
<td>.04</td>
</tr>
<tr>
<td>Sadness</td>
<td>.19</td>
<td>.05</td>
<td>.31</td>
<td>.06</td>
<td>.59</td>
<td>.05</td>
<td>.78</td>
<td>.05</td>
</tr>
</tbody>
</table>
Figure 4. Aftereffect magnitude (measured as shift in proportion of target responses from baseline to post-adaptation) for average-referenced and neutral-referenced trajectories. Bars indicate standard error. Aftereffects are significantly greater than zero.

For each participant we found the average aftereffect size for the average-reference and the neutral-reference trajectories, collapsing across expression (Fig 4). Aftereffects were larger for the average-referenced than the neutral-referenced trajectories. Kolmogorov-Smirnov tests indicated that the aftereffect distributions did not deviate significantly from normality (average-referenced: $D(24) = .122, p > .05$; neutral-referenced: $D(24) = .133, p > .05$). A paired samples t-test confirmed that aftereffects were significantly larger for the average-referenced trajectories than neutral-referenced trajectories, $t(23) = 5.92, p < .001, r = .78$. 
Single-sample t-tests indicated that aftereffects in both conditions were significantly larger than zero (average-referenced: $t(23) = 7.92, p < .001, r = .86$, neutral-referenced: $t(23) = 5.73, p < .001, r = .77$).

For two of the expressions, fear and disgust, the neutral-referenced anti-expressions contained artifacts around the mouth (Figure 2). These artifacts make it difficult to interpret the contours of the lips, which could impair the effectiveness of these antiexpressions as adaptors. While we made no predictions about the aftereffects on particular expression trajectories, we examined the data divided by expression to determine whether these artifacts were driving the difference in aftereffect size between the neutral and average-referenced trajectories (Figure 5).

![Figure 5](image)

*Figure 5.* Mean aftereffect magnitude for each adaptor. Bars indicate standard error. Neutral-referenced fear and neutral-referenced disgust adaptors contain visual artifacts.

Inspecting Figure 5, it is apparent that the expressions which feature potentially problematic neutral-referenced adaptors (fear and disgust) show a
much greater difference between average and neutral aftereffects than the expressions which do not (anger and sadness). A repeated measures ANOVA revealed a significant main effect of reference expression \( (F(1, 23) = 35.03, p < .001, \eta_p^2 = .60) \), no significant main effect of target expression \( (F(3, 69) = 2.03, p = .118, \eta_p^2 = .08) \), and a significant interaction between reference expression and target expression \( (F(3, 69) = 15.31, p < .001, \eta_p^2 = .40) \). Paired samples t-tests confirmed that the differences between the average-referenced and neutral-referenced conditions were significant for fear \( (t(23) = 8.54, p < .001, d = 1.89) \) and disgust \( (t(23) = 4.91, p < .001, d = 1.18) \), but not for anger \( (t(23) = -.19, p = .853, d = 0.05) \) or sadness \( (t(23) = -1.54, p = .136, d = 0.32) \). One-sample t-tests indicated that all aftereffects were significantly larger than zero for all trajectories except neutral-referenced fear and neutral-referenced disgust (Table 3).

Table 3: One-sample t-tests comparing size of aftereffect on each trajectory to zero.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Expression</th>
<th>( t(23) )</th>
<th>( p )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Anger</td>
<td>4.59</td>
<td>&lt; .001</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>8.45</td>
<td>&lt; .001</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>Disgust</td>
<td>5.51</td>
<td>&lt; .001</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>Sadness</td>
<td>2.69</td>
<td>.013</td>
<td>0.55</td>
</tr>
<tr>
<td>Neutral</td>
<td>Anger</td>
<td>5.90</td>
<td>&lt; .001</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>1.98</td>
<td>.060</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>Disgust</td>
<td>0.91</td>
<td>.375</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Sadness</td>
<td>4.25</td>
<td>&lt; .001</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Discussion

The main pattern of our results indicates that the average expression lies closer to the center of expression-space than the neutral expression. Adaptation to anti-expressions on trajectories that passed through the average expression produced larger aftereffects than adaptation to anti-expressions on neutral-referenced trajectories.

However, the difference seems to be largely based in the fear and disgust expressions, which are the two expressions that produced neutral-referenced anti-expressions with noticeable artifacts around the mouth area. Adaptation to neutral-referenced anti-fear and neutral-referenced anti-disgust produced almost no aftereffects in this study. It is possible that these artifacts prevented adaptation from producing an aftereffect in this condition. We consider this explanation unlikely to be true. The anti-expressions with artifacts appear odd, but are still clearly face-like. In identity aftereffect studies, extremely distorted caricatures have still been found to produce strong aftereffects (McKone, Jeffery, Boeing, Clifford, & Rhodes, 2014). The same also seems to hold for expression; Juricevic and Webster (2012) created anti-expressions relative to a neutral expression, resulting in artifacts around the mouth, but adapting to these anti-expressions still produced significant aftereffects.

Rather, the small aftereffects for these two anti-expressions may be related to a low baseline response rate. Participants rarely judged the neutral expression to be fearful or disgusted even in the absence of adaptation, and were much more likely to see it as sad (Table 2). Adaptation to anti-fear or anti-disgust does not seem to have shifted perception of the neutral expression sufficiently to make it appear more fearful or disgusted than sad. It is possible that adapting to these two
anti-expressions produced aftereffects, but that testing with the 0% reference face cannot measure them.

The problem encountered above points to a more general problem with our design. We are trying to measure and compare aftereffects by taking responses to a test stimulus at a single location on the expression trajectory (0%). This approach has been applied successfully in previous research. For instance, the near-far paradigm used to determine whether expression is opponent-coded compares aftereffects at 0% in two adaptation conditions: after adapting to a 50% anti-expression or a 100% anti-expression (Burton et al., 2013; Skinner & Benton, 2010, 2012). Aftereffect size has also been measured at single test level across a number other face aftereffect studies (e.g. Jeffery et al., 2011; Leopold, Rhodes, Müller, & Jeffery, 2005; Pond et al., 2013; Rhodes, Jeffery, Clifford, & Leopold, 2007; Rhodes et al., 2004; Robbins, McKone, & Edwards, 2007; Rutherford et al., 2008; Skinner & Benton, 2010, 2012). However, when using this approach it is important to consider that the effect of adaptation on responses at a single test level will depend on where that level lies on a participant’s response curve. If the test level is near the steepest part of the curve at baseline, then an adaptation aftereffect that produces a sizeable shift in the response curve will substantially change responses at that test level (Figure 6a). If the test level sits at a flatter portion of the curve, where responses are near floor or ceiling, that same shift in the curve will give a much smaller change in responses at that test level (Figure 6b).

This fact is not a problem for all experimental paradigms. For instance, in the near-far paradigm, we test at the same level (the same point on exactly the same response curve) in both adaptation conditions. If one adaptor produces a bigger change in responses than another, we know that the difference is because
that adaptor shifts the participant’s response curve more than the other adaptor. We only have to be careful to choose a test level that is on a steep enough part of the curve that responses will change following adaptation.

Figure 6. Demonstration of a potential problem with measuring aftereffects at a single test level. Solid line represents a baseline response curve; dashed line represents response curve after adaptation. Red line indicates test level. In both cases, adaptation shifts the entire curve by the same amount. However, depending on where our test level lies on the curve, the effect of adaptation on responses will vary. If we only measured responses at this one point, it would appear that the effect of adaptation is larger in A than in B.
Our paradigm in this experiment is different to the near-far paradigm. We want to compare the aftereffects produced by adaptation on two different test trajectories. Unfortunately, in this case testing at a single level is not appropriate. Without plotting out the full response curves, we cannot be sure that the 0% test level on the neutral-fear trajectory lies at the same point on a participant’s response curve as the 0% test level on the average-fear trajectory. This means that we cannot simply compare ‘aftereffects’ measured at a single level between the two conditions, because any difference in the measured responses may not capture the relative shifts in response curves. Our experiment, as designed, cannot tell us whether aftereffects are larger on the neutral-referenced or average-referenced trajectories.

We would also like to note that this issue of design logic might still be a problem even if we knew in advance that our test levels were at equivalent points on the two response curves. For instance, we might check and find that a participant responded “anger” 50% of the time to both the 0% neutral and 0% average expressions. Even though this participant is at the mean of their response curve in each condition, the two curves might have different slopes. In this case, a shift of the same size in the mean of the two curves would still result in a different “aftereffect” measured at 0% (Figure 7).

We should also consider that it is possible that adaptation might change not only the mean but also the slope of the entire response curve (as appears to be the case in Burton et al., 2015). We might start out testing at exactly the same point on two response curves that have identical slopes, but if the slope of one curve changes more than the other following adaptation, then even if the overall shift in the two curves is the same, the “aftereffect” measured at our test level will be bigger for one condition than the other (Figure 8).
Figure 7. Another scenario in which measuring an aftereffect at a single test level may be misleading. Again, solid line represents a baseline response curve, dashed line represents response curve after adaptation. Adaptation shifts the response curve over by the same amount in each condition, and in both cases the test level lies at the mean of the curve. However, when measuring at a single test level (red line) the effect of adaptation appears smaller when the slope of the curve is shallower.
Figure 8. A scenario in which the baseline response curve is exactly the same in both conditions at baseline (solid line), but in one condition the slope of the curve changes following adaptation (dashed line). In both conditions the overall effect of adaptation is the same (the mean shifts by the same amount, and the total effect on responses across all levels is equal), but at our single test level (red line) we will measure a larger change in responses for one condition than the other.

To summarise, when deciding how to measure an aftereffect it is important to consider whether we want to compare two aftereffects that are being produced relative to the same baseline response curve, or two aftereffects produced relative to different baseline response curves. If the two aftereffects are
produced relative to the same baseline response curve (i.e. the test stimuli are identical and the response task is the same in both conditions), and adaptation affects the slope of the response curve in the same way in both conditions, then measuring responses at a single test level is appropriate. If the two aftereffects are produced relative to different baseline response curves (i.e. the test stimuli are different and/or the response tasks are different between conditions), or adaptation in the two conditions affects the slope of the response curve in different ways, then a single test level is likely to be inappropriate, and may result in misleading data.

We aimed to resolve this problem in Chapter 4 by repeating the experiment, but using a wider range of test levels. We measured the aftereffect by fitting curves to the baseline and post-adaptation data and calculating the difference between the 50% thresholds of the two curves. The difference between the thresholds represents the overall shift in the response curve, making it a more appropriate measure for comparing aftereffects between the two conditions. When we used this measure of the aftereffect, we found significant aftereffects for both the neutral-referenced disgust and neutral-referenced fear trajectories (the trajectories for which we found no aftereffect when testing only at 0%). Examination of the functions fit to those data confirms that there was an overall shift in the response curve for neutral-referenced disgust and fear, but that at 0% responses were near floor and do not reflect that shift. When we measured the aftereffect as a shift in the 50% thresholds we found no overall difference in the size of aftereffects between neutral-referenced and average-referenced trajectories (see Chapter 4 for more details).

In the 0% test experiment reported here, we observed that the baseline response rates were not at chance for either the 0% average or 0% neutral faces
We explored the baseline response rates for the 0% expressions in the second (Chapter 4) experiment to check if this observation was replicated in the new dataset. We noticed that the baseline response rates differed between the two experiments. We report the baseline data from the second experiment in Table 4, and repeat the baseline data from the first experiment for comparison.

Table 4: Baseline response rates from the experiment reported here (Experiment 1) and from the 0% test levels in the Chapter 4 experiment (Experiment 2).
Response rates are the proportion of trials in which the average and neutral expressions were labeled as each of the target expressions with no adaptation.

<table>
<thead>
<tr>
<th></th>
<th>0% Average Expression</th>
<th>0% Neutral expression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experiment 1</td>
<td>Experiment 2</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SE$</td>
</tr>
<tr>
<td>Anger</td>
<td>.16</td>
<td>.03</td>
</tr>
<tr>
<td>Fear</td>
<td>.39</td>
<td>.05</td>
</tr>
<tr>
<td>Disgust</td>
<td>.26</td>
<td>.04</td>
</tr>
<tr>
<td>Sadness</td>
<td>.19</td>
<td>.05</td>
</tr>
</tbody>
</table>

In particular, the proportion of average expressions labeled “anger” decreased considerably from the first experiment to the second, with a corresponding increase in the proportion labeled “disgust”. Participants performed the same baseline task in both experiments, viewing a test expression and then choosing between the same four responses to label it. However, this task was performed in different contexts in the two experiments. One difference is that in the original version of the task participants saw a grey oval ‘adaptor’ before the test face, which was not present in the revised version. However, it is not clear
how this difference might have contributed to the change in baseline response levels. Additionally, in the original version test faces were mostly the 0% reference faces (with infrequent 60% target faces), while in the revised version the test faces covered a range of expression strengths (from -20% to 80%) from each of the eight trajectories. Participants therefore viewed many more instances of the target expressions at various strengths while completing the baseline trials of the revised version compared to the original version. These expressions may have affected participants’ response criteria by providing reference points against which to compare the 0% test face. Whatever the cause, this effect suggests that it is important to consider the context in which participants are performing a task, particularly if this measurement will serve as a baseline against which other conditions will be compared.

In this experiment we aimed to determine whether the average or neutral expression lies closer to the center of expression space. We compared the size of the aftereffects produced by adapting on average-referenced trajectories to aftereffects produced by adapting on neutral-referenced trajectories. We used a common method of measuring the aftereffect – testing responses at a single test level, and observing how these responses change with adaptation. We found larger aftereffects for average-referenced adaptors than neutral-referenced adaptors, but this effect was driven by floor effects for two trajectories. The 0% neutral test face was almost never labeled as fearful or disgusted at baseline, and adaptation to neutral-referenced anti-fear or neutral-referenced anti-disgust did not increase participants’ tendency to respond with these labels. When we measured responses across a wider range of test levels in a subsequent experiment, we found aftereffects on both of these trajectories. These findings highlight a wider methodological issue: When comparing aftereffects between conditions,
measuring the aftereffect at a single test level may give misleading results. This method of measuring the aftereffect can be appropriate in certain circumstances, and can cut down on testing time. However, in other circumstances it is more appropriate to plot a wider range of the response function and measure the aftereffect as a shift in that function.
References:


McKone, E., Jeffery, L., Boeing, A., Clifford, C. W. G., & Rhodes, G. (2014). Face identity aftereffects increase monotonically with adaptor extremity
over, but not beyond, the range of natural faces. Vision research, 98(0), 1-13.


