University of Nevada, Reno

# Explorations of Face Space

A dissertation submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy in

Psychology

by

Sean O'Neil

Dr. Michael A. Webster/Dissertation Advisor

May, 2014

University of Nevada, Reno
Statewide • Worldwide

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

**SEAN O'NEIL**

entitled

**Explorations Of Face Space**

be accepted in partial fulfillment of the
requirements for the degree of

**DOCTOR OF PHILOSOPHY**

Michael A. Webster, Ph. D., Advisor

Michael A. Crognale, Ph. D., Committee Member

Gideon P. Caplovitz, Ph. D., Committee Member

George Bebis, Ph. D., Committee Member

Thomas J. Nickles, Ph. D., Graduate School Representative

Marsha H. Read, Ph. D., Dean, Graduate School

May, 2014

# Abstract

According to a norm-based "face space" neural representation, faces are arranged by how far they deviate from a norm or average face, with faces of increasing distinctiveness represented by their distance from the norm. This is supported by evidence from adaptation, development, binocular rivalry, and neuroimaging. A series of experiments examines the norm-based structure of face space. First, adapting to young or old faces caused perceived age aftereffects. When middle age faces were adapted, no shift is shown. Both of these results are consistent with representation relative to a norm based on comparisons of pooled neuronal activity. Next, rivalrous neutral and distorted faces were presented binocularly, creating a constantly changing percept. Distorted faces were clearly dominant in initial experiments, but this discrepancy seems to be partially controlled by the distinct face having a greater pull in categorization. Increasing the types of response choices or limiting fusion decreased the bias, with evidence that observers were often seeing a partially distorted face. Finally, other experiments examined our ability to recognize angular and radial changes in a face space representation. Our ability to consciously access our own face space appears lacking, as well as our ability to equate perceived intensities between faces, especially with faces distant in face space. The evidence presented demonstrates that face space is consistent with a norm-based layout, and that this model might accurately be represented within a relative coordinate system similar to color and other models.

# Acknowledgements

I would like to thank my advisor, Michael A. Webster, for his mentoring and support throughout graduate school. I would also like to thank my collaborators for assistance and advice in Experiment I: Amy Mac and Gillian Rhodes, and in Experiment II: Gideon P. Caplovitz.

Research was supported by NIH grant EY-10834.

Specific face stimuli in Experiment I were from Center for Vital Longevity Face Database, University of Texas at Dallas (Minear & Park, 2004) and in Experiment II from the Radboud Faces Database (Langner et al., 2010). Experiment III used FaceGen generated stimuli (Singular Inversions).

# Table of Contents

Abstract ........ i

Acknowledgements ........ ii

Table of Contents ........ iii

List of Tables ........ v

List of Figures ........ vi

**Introduction** ........ 1

Face Space ........ 1

Representation of Faces Relative to the Norm ........ 3

Research Questions ........ 9

I. Perceived age of faces ........ 13

Abstract ........ 13

Introduction ........ 13

Method ........ 18

Results ........ 19

Discussion ........ 22

II. Sibling Rivalry ........ 24

Abstract ........ 24

Introduction ........ 24

Method ........ 27

Results ........ 34

Discussion ........ 37

III. Perceiving relationships within face space ........ 43

Abstract ........ 43

Introduction ........ 44

Method ........ 48

Results ........ 53

Experiment 2: Intensity Matching ........ 59

Method........................................................................................................... 59
**Results** ........................................................................................................... 60
Experiment 3: Averaging Faces...................................................................................... 64
**Method** ........................................................................................................... 64
**Results** ........................................................................................................... 65
Discussion ........................................................................................................... 67
IV. General Discussion ........................................................................................................... 72
References ........................................................................................................... 80

# List of Tables

Table 1: Number of observers that set matches significantly different from 180° opposite predictions ($p < 0.05$) and percentages of observers from the total. .................................... 58

# List of Figures

**All figure error bars represent standard error.**

Figure 1: Face space representation of the happy vs. angry plane ....................................... 6
Figure 2: Norm-based (top) vs. exemplar-based (bottom) accounts of face coding and face aftereffects ........................................................................................................ 15
Figure 3: (a) Age estimates before (circles) or after adapting to young (inverted triangles), old (squares) or middle-aged faces (diamonds), as a function of actual age. Settings are replotted as a function of pre-adapt perceived age in (b) Young adapt (c) Old adapt (d) Middle age adapt ....................................................................................... 21
Figure 4: Example of the stimuli used in the Sibling Rivalry experiment .......................... 28
Figure 5: Possible responses in Sibling Rivalry Experiment 3............................................ 33
Figure 6: Stimuli used in Experiment 4.................................................................................. 34
Figure 7: Binocular Rivalry Experiment 1 results................................................................. 35
Figure 8: Binocular Rivalry Experiment 2 results................................................................. 36
Figure 9: Binocular Rivalry Experiment 3 results................................................................. 39
Figure 10: Binocular Rivalry Experiment 4 results............................................................... 40
Figure 11: Face space representation of angry, happy, and their respective anti-faces .... 51
Figure 12: Example of two observers' settings using reference faces in the fear/happy continuum ................................................................................................................ 54
Figure 13: Opposite face settings (y axis) as a function of reference face (x axis).............. 55
Figure 14: Dark blue: prediction line for 180° shape opposite settings. Other colors: settings made by 4 observers on a fear/happy pair.............................................................. 57
Figure 15: Averages for between-expression intensity matches ......................................... 62
Figure 16: Settings for within-expression intensity matches ............................................... 63
Figure 17: Intensity settings (circles) as an average of corresponding reference faces presented at fixed intensities (same-color divided squares)................................................ 66

# Introduction

## Face Space

Simply by interacting in a social environment, we are constantly in contact with other faces, known and novel, and must constantly identify and classify them or interpret their current affect or emotional state. Neurally, face perception draws on low-level inputs like the color, spatial frequency, and contrast of a face. Yet more importantly, a number of lines of evidence suggest that face perception may also critically depend on high-level processes (e.g. configural coding or holistic processing) and specialized processing modules such as the fusiform face area. Face perception and its neural correlates have seen an explosion of interest in the last decade: the number of articles published per year has increased from 458 in 2002 to 1,603 in 2012 (Web of Science). This has led to a greater understanding of face perception in many ways, from how we recognize the identity of a face, to our perception of expressions, to our perception of other aspects such as race or gender. Many of these aspects can be integrated into a common theoretical framework in which an individual face is represented in a multidimensional space, with each face coded by how it differs relative to an average or prototype face (a norm-based code). In analogy to the many color spaces used in vision science, this has been termed a "face space" (Valentine, 1991), and represents each face by the degree and direction to which it differs from a perceptually average face, or one that is an average of many genders, ethnicities, identities, and expressions. This dissertation explores the extent to which a metrical "space" of this kind provides a reasonable metaphor for how faces are actually represented in the visual system.

In face space, one way of coding faces based upon different categories, whether they are expressions, identities, etc., is to represent them as vectors relative to the prototype face. This prototype is a continuously updated face, representing the average of faces recently encountered (Principe & Langlois, 2012; Rhodes, Maloney, Turner, & Ewing, 2007). This is analogous to how hue is represented in many color spaces such as DKL space (Derrington, Krauskopf, & Lennie, 1984). In this type of space, one aspect of color can be independently manipulated without affecting another aspect, e.g. relative levels of red vs. green can be adjusted without affecting the blue vs. yellow or luminance components. Colors that are orthogonal to each other in color space and thus can be adjusted independently, like red and blue, can be perceptually mixed, while "impossible" mixes like red-green represent opposites or mutually exclusive sensations along the same axis. Crucially, mixing direct opposites would create a neutral gray and not a red/green. Other color space models, like the Munsell color system (Munsell, 1905), also have three dimensions or attributes of color that are represented independently (e.g. hue, chroma, and value), with hue represented by angles surrounding cenral gray. In general then, color spaces can be sorted into two main divisions depending on whether they represent colors by their *absolute* coordinates (L, M, S cone excitations) or by their positions *relative* to a central, neutral (achromatic) position (e.g. separate numbers for hue, saturation, and brightness or luminance).

The layout and nature of relationships in face space still remain unclear. But as the “redness” of a vivid red vs. a dull red can be represented by the length of a vector from some neutral point, current evidence supports representing faces in a similar manner. Here, hue can be seen as analogous to the *type* of facial category. In face space, there are multiple spaces that might exist for different expressions, identities, etc., with the commonality that at the center of this face space is a neutral face and the more peripherally that faces are represented, the less they appear like the neutral face (Figure 1) (de Fockert & Gautrey, 2013). Intermediate between “pure” expressions are those which ambiguously resemble two expressions simultaneously to some degree, similar to various shades of purple which could be characterized by their relative similarity to red or blue. And opposite colors, or opposite faces, if averaged, will tend towards gray or a neutral face instead of combining. While the midpoint between the two might be difficult to classify as belonging to either category, we might potentially classify it by which pure expression it *best resembles*. Additionally, saturation in color models can be analogous to the *strength or intensity* of a facial attribute; for example a face that is somewhat happy will have a subtler smile.

## Representation of Faces Relative to the Norm

Vital to the circular face space model that has distinct faces surrounding a neutral center, is that along with intensity decreasing as faces are represented closer to the center, a face immediately opposite of the neutral point on that same linear vector is a face that has an opposite expression. This brings up an important question: what face is the opposite of an expression such as happiness? One possibility is that our face space represents the opposite as its emotional or conceptual opposite, e.g. sadness is the opposite of happiness. However, this choice is not as obvious for other expressions; another possible answer is that anger is the antithesis of happiness. Yet another possibility is that we represent the "shape opposite" of the expression, a characteristic which has been examined for attributes such as identity (Jeffery et al., 2011). These opposite faces are termed "anti-faces" (Leopold, O'Toole, Vetter, & Blanz, 2001). Anti-faces could provide powerful tools for showing the arrangement of face space. Adaptation can be used as a tool to investigate anti-face aftereffects, e.g. adaptation to a happy face creates an anti-happy percept, following adaptation models in color and other modalities. Moreover, finding the opposite relationship, that adapting to anti-happy causes a happy aftereffect would provide important conceptual information about expression face space and its symmetric, opponent arrangement (Rhodes & Jeffery, 2006).

The representation of faces in face space has been applied to two distinct models. The first represents the location of faces by their absolute *coordinates* in face space, and neurally represents faces by how closely they fit a series of face category specific neurons that are narrowly tuned (exemplar-based coding) (Valentine & Endo, 1992; Valentine, 1991). A face's percept then derives from which neurons are active, e.g. if neurons selective for 45-year old faces are firing the most, this translates to a face being perceived as around 45 years. This model is analogous to some representations of low-level vision, such as spatial frequency (Webster & MacLeod, 2011). The second model more closely resembles physiological models of color vision and represents faces by their degree of deviation from a norm, relying on the comparison of broadband *pooled* outputs of two *populations* of neurons (norm-based coding). The ratio of activation is then translated to the position in face space; a less balanced ratio notes a more extreme face. The existence of a neutral or average face is central to norm-based models (analogous to achromatic gray in color vision), as it provides a reference point for other faces to define their differences from it and from each other (Figure 1, or see (Jeffery et al., 2011) for several descriptive examples).

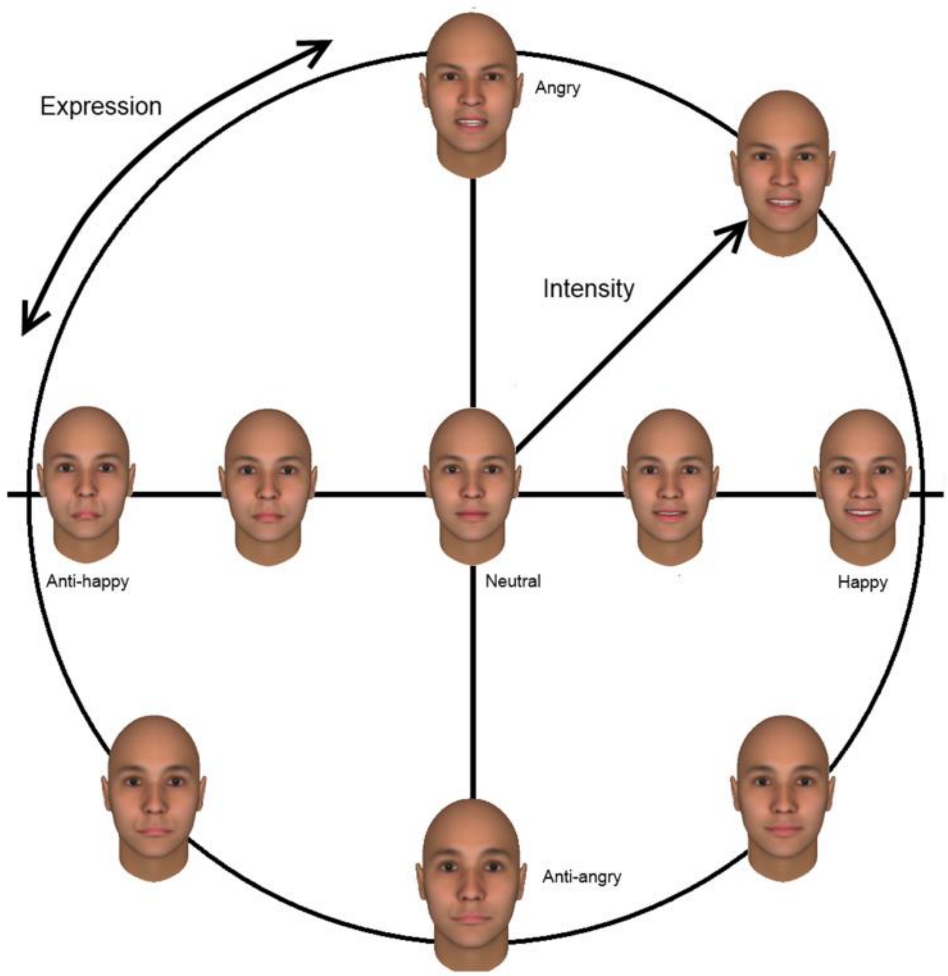


**Figure 1: Face space representation of the happy vs. angry plane. Intermediate faces on the happy vs. anti-happy intensity axis and expression axis are shown.**

Adaptation research has generally supported a norm-based model. Aside from aftereffects appearing like the face directly opposite the norm, the intensity in face space seems to have an effect. More extreme caricatures provide stronger aftereffects than a weaker face, consistent with the faces being relative to a norm, and aftereffects are often selective for the same face axis but not for other expressions (Juricevic & Webster, 2012; Oruç & Barton, 2011; Webster, Kaping, Mizokami, & Duhamel, 2004). There are other studies that suggest that the norm-based model has greater explanatory power. This is evidenced in the fact that, e.g. adapting to a full smile provides a larger aftereffect than a half-smile, while adapting to a neutral face will not bias percepts in any particular direction, consistent with it being a norm. With an exemplar-based model, adapting to a strong expression predicts even smaller aftereffects, as the adapting stimulus is farther from the region of overlap for a neuron selective for a specific expression intensity (Robbins, McKone, & Edwards, 2007). What support exists for exemplar-based models seems to involve specific mechanisms such as gaze direction (Calder, Jenkins, Cassel, & Clifford, 2008). But for the majority of aspects of face perception, such as expression, norm-based coding remains the better explanation.

There are several effects which provide further support for a norm-based model of face space. The first are caricature effects. They represent a face that is farther from the norm than even the veridical face, and are created by extrapolating beyond the veridical face using similar but opposite shape manipulations as those needed to create an anti-face. Caricatures appear exaggerated e.g. a male face may be caricatured as extremely male, or an identity's distinguishing features will become exaggerated. Caricatured faces appear to be easier to identify and are viewed as more distinctive, when compared to anti-faces (Lee, Byatt, & Rhodes, 2000), and it seems that adaptation to caricatures facilitates recognition of similar faces more than with less distinct adaptors, fitting a norm model (Rodríguez, Bortfeld, Rudomín, Hernández, & Gutiérrez-Osuna, 2009). With regards to acquisition of a face space, it appears that while experience may hone conceptual face space, its existence seems to be innate or acquired early. Developmental studies with normal children as young as 4 years have found that their face perception is similar to adults, with adaptation aftereffects consistent with a norm (Jeffery et al., 2010, 2011). Evidence is also shown outside the realm of psychophysics and neurotypical observers. Adult individuals with prosopagnosia, an impairment of face perception, perform worse than neurotypicals in tasks involving identity recognition. Yet surprisingly, individuals with congenital prosopagnosia perform similarly when doing non-identification tasks with caricatures, equal blends between average and veridical faces (anti-caricature), and anti-faces. This suggests norm-based processing even in prosopagnosia, even if they cannot identify or recognize these faces. However, impairment in face space processing was greater in an individual with acquired (brain damage) prosopagnosia (Nishimura, Doyle, Humphreys, & Behrmann, 2010).

Neuroimaging research also supports similar results. Experiments by Loffer et al. (2005) used functional magnetic resonance imaging (fMRI) to looked at adaptation effects on the BOLD signal in the fusiform face area. They found that greater BOLD signal strength in the fusiform face area was correlated with faces that were farther from the average face or very distinct. In addition, after adapting to a face of a particular identity, BOLD signals were attenuated (i.e. “fatigued” due to overexposure) if a face of the same identity was viewed, even if it varied in degree of distinctiveness (how caricatured it was). Conversely, the signal was not attenuated if a different face was viewed. Even in non-human primates, evidence shows neurons that are preferentially selective for faces tend to be tuned around the average face (Leopold, Bondar, & Giese, 2006). The inferior temporal cortex in macaque monkeys has cells which respond more strongly to distinct or caricatured human faces than average ones, consistent with a norm-based comparison between average and distinctive faces (Leopold et al., 2006). These experimental results suggest a neural representation of a norm that is contrasted with other faces, which are treated based on how far they are from the norm.

While these results are consistent with a norm-based code for faces, they do not necessarily imply that different faces are mapped in a way that preserves the metrical relationships between them. For example, the visual system might represent each face by how distinctive it is relative to the average, while not representing how one form of distinctiveness differs from another. By this account, different identities might be coded as qualitatively different “objects” (like apples and oranges), and observers may therefore not have a way to estimate the quantitative differences (vector angles) between them.

## Research Questions

The following experiments are designed to test face space-based processing across multiple and varied tasks. The concept of a face space that encompasses expression, gender, race, identity, etc. is over 20 years old (Valentine, 1991) but there are still many questions about how it is represented (norm or exemplar?). For example, it is entirely unknown to what extent we might have conscious access to this representation, and whether the wide variety of phenomena within face perception can be accounted for by a common model or by models that are functionally similar.

Three main experiments address different aspects of face perception and how they relate to face space, by testing whether they are processed in an opponent system that relies on faces being represented in reference to a central norm. The first set of experiments looks at how we process age when viewing faces. We are constantly adapting towards the average of a vast number of faces as we encounter them in our environment. Consequently, we create an internal norm which is constantly being updated. This experiment uses adaptation as a tool to highlight how the norm might be updated through continuous exposure to face stimuli specifically by examining how age judgments are affected by adaptation across a wide range of ages. The experiment focuses on age because, unlike all other facial attributes, observers are very good at guessing the age of a face, and thus "quantifying" the full spectrum of variations along the dimension. This provides a unique and powerful tool for quantifying the actual pattern of face aftereffects, or how adaptation to one age affects all others. For example, this can be used to test whether the aftereffects are "local," and thus only influence the perception of faces that are physically similar in age to the adapting face; or are more "global," so that adapting to one age produces roughly constant effects on the perception of all ages. In turn, the pattern of these aftereffects can be used to distinguish between norm-based vs. exemplar-based models of the encoding of facial age.

The second set of experiments use the method of binocular rivalry, wherein two images are displayed to each eye but only one is usually perceived at any given time due to rivalry between the two competing inputs. Observers' perceptions were probed while they viewed normal and distorted faces in conditions of binocular rivalry. Previous studies have found that during rivalry, perception is dominated by the “stronger” stimulus. This includes stimuli that are stronger because they have a higher contrast (Mueller & Blake, 1989), and this experiment tests whether more distinctive faces dominate in rivalry because they have a higher “contrast” within face space. To vary distinctiveness we used faces that were configurally distorted, since these change the relationships between different facial features (e.g. separation between the eyes) without changing the low-level characteristics of the images (Yamashita, Hardy, De Valois, & Webster, 2005). The distorted face was hypothesized to dominate in rivalry because it represents a stronger “saturation” of vector length in face space.

In the third set of experiments, we assess the observers’ ability to access their own face space, by determining whether they have any intuition about the metrical relationships between different expressions, and what motivates their selection (e.g. whether they are based on conceptual knowledge about opposing emotions or on perceptual knowledge about visual facial configurations). Expressions have the advantage that they are few in number and well defined (Ekman, Sorenson, & Friesen, 1969), so that the primary axes of the space are plausibly known. The experiments will specifically test whether observers represent expressions in terms of actual spatial coordinates, by testing whether they can tell when two expressions fall on opposite sides of the space. Further experiments will look at more complex indications of an internally-referenced face space. As faces can also be represented by their distance from the norm in face space as intensity, the relationship with and interactions between the form of expression and its intensity are examined. Collectively, these experiments will help to reveal the coding schemes that the visual system uses for faces, and whether a "space" and face coordinates do in fact characterize how we perceive faces. The results of these experiments will also provide important evidence for whether very different stimulus domains (e.g. color and faces) are represented in the visual system using similar or distinct strategies, and thus are relevant to the general issue of how information is encoded in the brain and the extent to which the brain relies on common coding principles.

# I. Perceived age of faces

**Abstract**

Adaptation aftereffects have been widely used as a tool to probe how experience shapes the visual encoding of faces, but the actual changes in facial appearance produced by adaptation and the neural mechanisms these imply remain poorly characterized. We explored how adaptation alters the perceived age of the face, a fundamental attribute of face perception for which the full range of natural variations can be reliably scaled by the observer. Observers guessed the ages of faces that ranged from 18-89 years old, before or after adapting to different faces of younger, older, or middle-aged adults. Prior exposure to the young or old group induced opposite linear shifts in perceived age that were independent of the model's age, while the middle-aged adaptors induced no aftereffects. This pattern suggests that adaptation leads to a simple and uniform renormalization of age perception, and is consistent with a norm-based neural code for facial age.

## Introduction

Adaptation aftereffects have come to play a prominent role in studies of the visual representation of faces (Rhodes & Leopold, 2011; Webster & MacLeod, 2011). After viewing a distorted face (e.g. with eyes pulled apart), a normal face looks distorted in the opposite direction (e.g. eyes too close-set) (Webster & MacLin, 1999). Robust aftereffects occur for "distortions" along which faces naturally vary, and thus can affect many of the cognitive and social judgments made about the face. These include adaptation-induced biases in perceived identity, gender, ethnicity, or expression, as well as attributes such as how attractive or trustworthy the individual appears (Engell, Todorov, & Haxby, 2010; Leopold et al., 2001; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Webster et al., 2004).

A number of studies have focused on the specific form of these aftereffects in order to infer the coding scheme underlying face perception. Two prominent alternatives for these schemes have been a relative or norm-based code, in which individual faces are represented by how they deviate from a prototype or average face; and an absolute or exemplar code, in which faces are coded by template matching (Figure 2) (Valentine, 1991). These differ in whether the average face has a special status, and predict different patterns of adaptation (Rhodes & Leopold, 2011; Webster & MacLeod, 2011). In norm-based codes, adaptation to an individual face renormalizes face percepts so that the adapting face appears less distinctive. This tends to recenter "face space" nearer to the adapting level, and predicts no aftereffects when adapting to the average face (since it is already at the norm). Thus in this model the adapting face should induce similar biases throughout the dimension (e.g. after adapting to a male face, ***all*** faces along the gender continuum should appear more female). In exemplar codes, adaptation should instead lead to more localized aftereffects by reducing sensitivity in channels tuned to the adapt face. This predicts no change in the appearance of the adapting face itself, but that other faces will look less like the adaptor. Thus aftereffects should be in opposite directions for faces on opposite sides of the

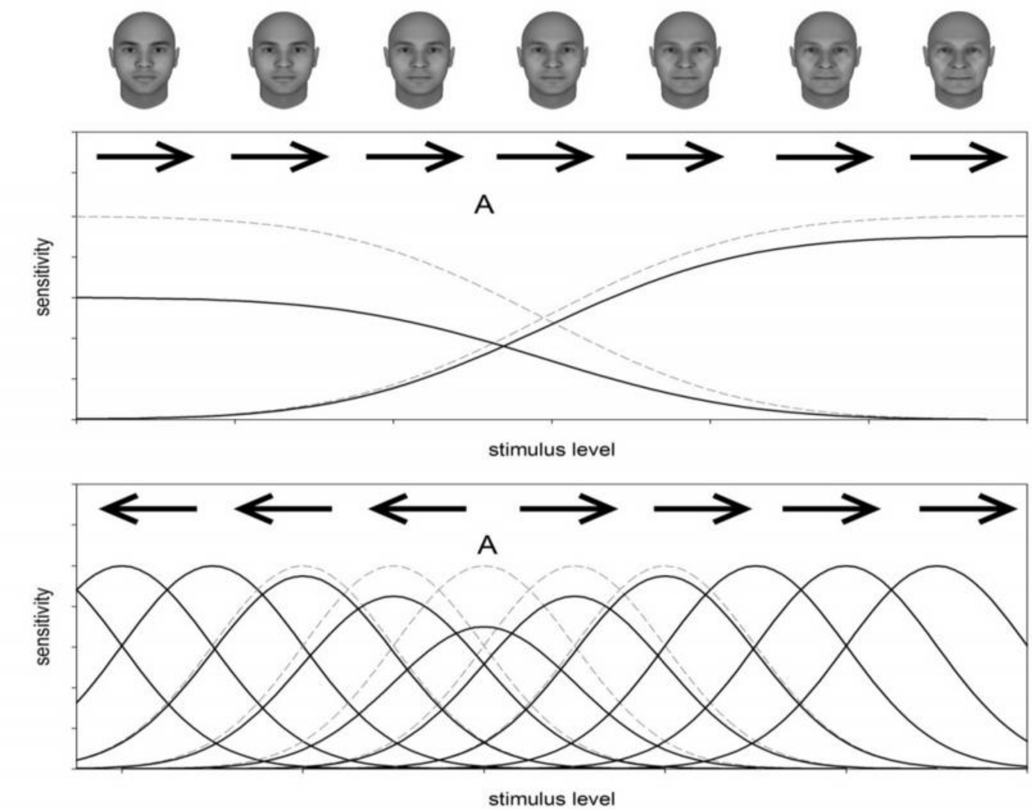


**Figure 2: Norm-based (top) vs. exemplar-based (bottom) accounts of face coding and face aftereffects. These differ in whether the levels of the face (e.g. age) are encoded by the relative differences in two broadly tuned mechanisms or by the peak response within channels narrowly tuned to different levels. Channel sensitivities are depicted before (dashed), or after (solid lines) adapting to stimulus level A. Arrows show the predicted direction of aftereffects that adaptation to level A induces in different age levels.**

adaptor. This model also differs in that the average face is coded by a template in the same way as other faces and should thus lead to similar aftereffects.

Much evidence from adaptation has been mustered in support of a norm-based code. This includes asymmetries in the aftereffects between normal and distorted faces (Webster & MacLin, 1999), common directions of distortion aftereffects so that after adapting to an expanded face all faces – including the adapt face – appear less expanded

(Rhodes et al., 2003; Robbins et al., 2007; Storrs & Arnold, 2012) and stronger aftereffects between faces that lie on opposite sides of the average face (Rhodes & Jeffery, 2006), and for faces that are more distinctive and thus farther from the norm (Skinner & Benton, 2010; Susilo, McKone, & Edwards, 2010). However, the possibility that adaptation reflects a general norm-based coding scheme for all or even most aspects of the face has not gone unchallenged. First, some facial attributes such as viewpoint and gaze direction show patterns of adaptation that are more consistent with a multichannel code (Calder et al., 2008; Fang & He, 2005). Second, recent studies have argued against renormalization for gender aftereffects because the adapting gender itself may not become less distinctly male or female (Storrs & Arnold, 2012), and because the aftereffects may not increase monotonically with gender strength (Zhao, Seriès, Hancock, & Bednar, 2011). Finally, recent models have questioned the extent to which adaptation can in fact distinguish between norm-based and exemplar codes (Ross, Deroche, & Palmeri, 2013). Thus both the form of the aftereffects and their implications for the neural coding of faces remain actively debated.

One factor hindering a full understanding of face adaptation is that aftereffects are usually only assessed for a limited range of target faces, and often only those near the prototype or boundary between face categories (e.g. male vs. female or angry vs. happy). This is in part because it is much easier for an observer to say *whether* a face is male or female, than *how* male or female it appears. However, from these restricted measures it is not possible to evaluate the ways in which adaptation is actually altering percepts of the facial dimension. Rating scales have been used to more fully capture the response changes, but have largely been limited to unnatural dimensions such as distortions, and only coarsely sample the dimension (Rhodes et al., 2003).

We took advantage of one attribute where fine metrical judgments about the face can be made – facial age. Age is one of the most salient characteristics of the face, and guessing someone's age is not only routine and intuitive but can be done with good accuracy for faces throughout the lifespan (Burt & Perrett, 1995; George & Hole, 2000). Thus age estimates provide a unique and sensitive psychometric scale for probing a fundamental dimension of face perception over the entire range of natural variation. Previous studies have shown that like other prominent facial dimensions, perceived age can be strongly biased by adaptation, and have shown that these aftereffects are driven by both the textural and to a lesser extent the shape changes that accompany aging (Lai, Oruç, & Barton, 2013; O'Neil & Webster, 2011; Schweinberger et al., 2010). They have also suggested that the adaptation to age at least partly reflects higher levels of visual coding(Lai, Oruç, & Barton, 2012). However, this work has again been limited to assessing how a young or old adapt faces alters the appearance of intermediate ages. We instead probed the effects of adapting to one age on the entire span of adult ages. This allowed us to assess specifically how adaptation alters the visual code for age, and thus to determine whether the code is norm-based or exemplar-based.

## Method

### *Observers*

Observers were 27 young adults at the University of Nevada. Nineteen adapted to young and old faces and 9 to middle-aged faces (one doing both). Observers gave informed consent and testing followed protocols approved by the university's Institutional Review Board.

### *Stimuli*

Stimuli consisted of digital color photographs of neutral-expression, Caucasian faces with recorded ages, available from the Center for Vital Longevity face database (Minear & Park, 2004). To reduce potential age cues from hair (Wright & Sladden, 2003), male images were clean shaven and all faces were cropped with an oval window (3:4 ratio). Eighty images were used as test faces and 30 as adapting faces. The test faces were relatively uniformly spaced from 18 to 89 years (M=50.54, SD=23.09) with similar numbers of males (38) and females (42). Adapt faces corresponded to a different set of 10 younger (M=34.3, SD=2.54) or older (M=65.0, SD=1.33) individuals, with 5 from each gender. A third group of 10 middle-aged faces was also used, based on the observers' ratings of age as described below.

### *Procedure*

Stimuli were presented on a CRT color monitor in a darkened room. During testing, the 80 test faces were displayed in random order for 500 msec, and observers had 4 sec after each to enter the estimated age using a handheld keypad. Observers were told only that all faces fell within the 2 digit range of 10-99 years. A practice trial with different faces was provided to familiarize them with the task. Settings were made in separate counterbalanced sessions in which a gray screen was shown before each face (no adapt), or with faces from a given adapting set displayed. During adaptation, the 10 adapt faces were first shown in a random sequence of 500 ms each for 2 minutes, and were repeated during the 5 sec between each test. To limit low-level aftereffects, adapt faces (subtending ~6.4° x 8.3°) were shown 1.5 times larger than test faces (~4.25° x 5.50°) and were jittered in position.

Observers tested with the middle-age adapt group first completed an additional session to estimate the age corresponding to the category boundary for old vs. young. The 80 test faces were shown in random order, but instead of estimating the specific age observers responded whether the model appeared “more young” or “more old.” A probit function was fit to the responses to determine the old/young boundary. This averaged 47 yrs (SD=6.7), and was used to define the middle adapting age.

## Results

As expected, observers were highly reliable at judging age from the face photos, though these judgments tended to compress the range of actual ages (Figure 3a). Physical age accounted for ~94% of the variance in the age estimates under each adapting condition. However, much of the residual variance was not because observers’ estimates were inconsistent but rather because some models did not look their real age.

This can be seen by replotting the post-adapt settings as a function of the pre-adapt settings (Figure 3b-d), which removes most of the scatter ($r^2$>0.98). Regression lines fitted to these relative settings reveal a surprisingly simple and systematic effect of adaptation on age percepts. Adapting to the young (Figure 3b) or old (Figure 3c) faces shifted the intercepts up or down by ~2.5 years, a bias that is small relative to the wide range of physical ages tested but is nevertheless highly significant (see Figure 3). In contrast, there was not a measurable effect on the slopes. Thus, the aftereffect is well approximated by a constant bias independent of physical age. After adapting to younger or older faces, *everyone* looked a few years older or younger, respectively.

This pattern indicates a renormalization of perceived age and thus a norm-based code. It is not compatible with simple versions of the exemplar model, which instead predict a steepened slope centered at the adapt level and aftereffects that are more localized to the vicinity of the adapting age. A further prediction of renormalization is that there should be no aftereffect if observers instead adapted to a “neutral” age, since this exposure would simply reinforce the norm. This was in fact the case - adapting to ages near the old/young boundary (“middle age”) did not produce a significant change in the mean settings or the slope relative to the pre-adapt settings (Figure 3d).

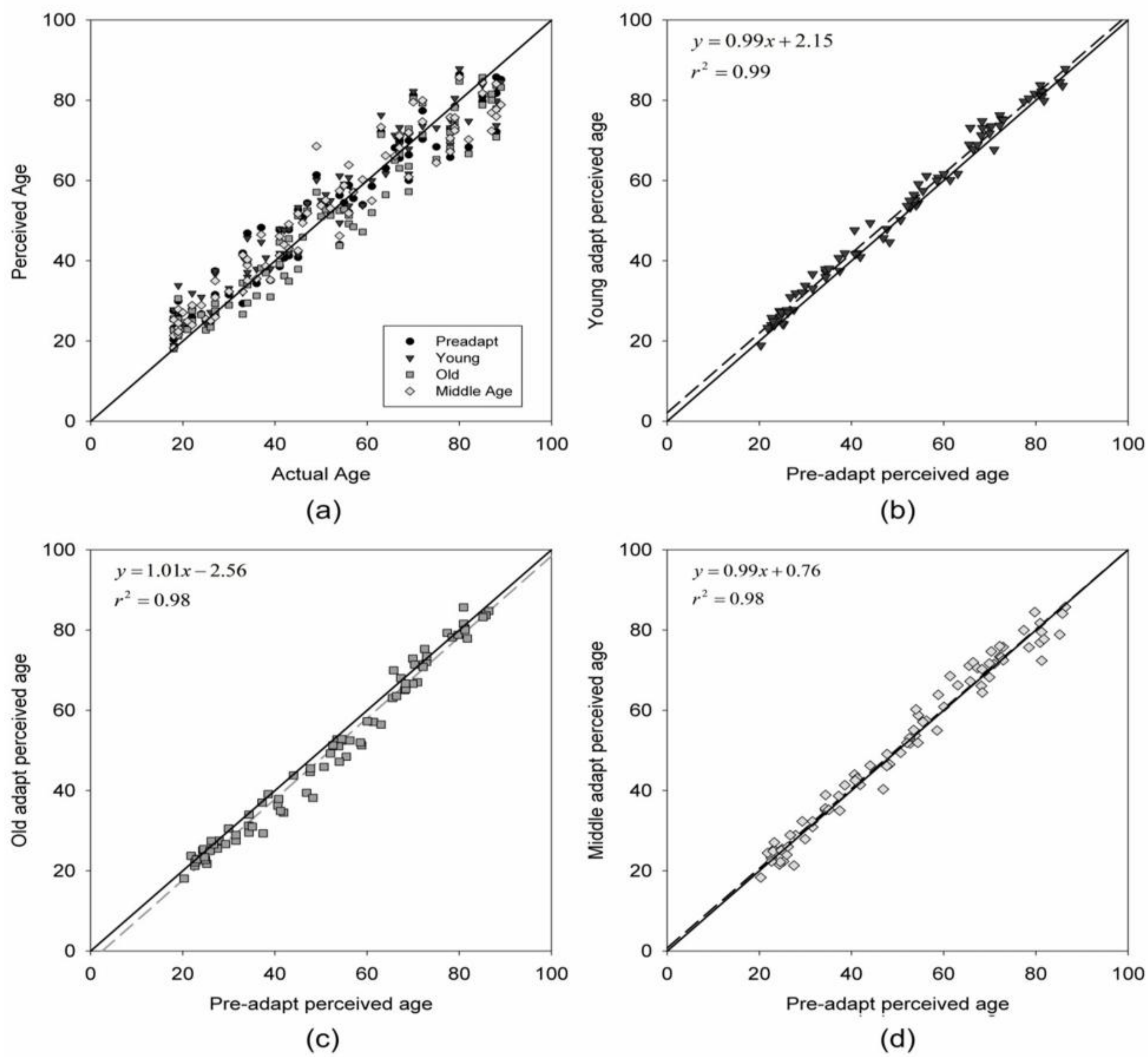


**Figure 3: (a) Age estimates before (circles) or after adapting to young (inverted triangles), old (squares) or middle-aged faces (diamonds), as a function of actual age. (b) Young adapt settings replotted as a function of pre-adapt perceived age; intercept = 2.15, t(78) = 3.25, p = .002; slope = 0.99, t(78) = 0.65, NS, (c) Old adapt settings as a function of pre-adapt; intercept = -2.56, t(78) = -2.97, p = .004; slope = 1.01, t(78) = 0.63, NS, (d) Middle age adapt vs. pre-adapt; intercept = 0.76, t(78) = 0.815, NS; slope = 0.99, t(78) = 1.30, NS. Lines show the linear regression for adapting conditions (dashed) and the unit diagonals (solid).**

## Discussion

By exploiting observers' natural expertise in evaluating age, our results reveal for the first time a full characterization of how adaptation alters a fundamental dimension of face perception. The aftereffects reflect simple linear shifts in perceived age – literally adding or subtracting a constant to the observers' estimates. Moreover, the adaptation appears to reset these estimates relative to faces that appeared neutral in age (i.e. "middle-aged"). Thus the results are well described as a uniform renormalization of facial age percepts so that the norm for age is adjusted toward the ages observers are currently exposed to. Measuring the aftereffects at multiple levels of the face dimension provides a strong test for this code. It also provides a much richer description of the actual impact of adaptation on face perception than previous measurements that only assess aftereffects around the norm or category boundary, which give little clue about how the full spectrum of faces might appear altered when we are exposed to a specific set of faces. For age these alterations again reflect simple recalibrations of the mean, with no measurable change in the variance, and are very reminiscent of the uniform shifts in color appearance with chromatic adaptation. Such parallels reinforce the idea that the visual system may often adopt similar strategies of norm-based coding in very different stimulus domains (Webster & MacLeod, 2011; Yovel & Belin, 2013), which we show includes the encoding of facial age.

It remains to be seen whether the adaptation pattern for age applies to other facial dimensions. The perception and adaptation of adult age is heavily dependent on face texture and thus may differ from the shape cues underlying other facial judgments (Lai et al., 2013; O'Neil & Webster, 2011). This difference may include the perception of age in children, for which structural changes in the face are much more pronounced (Enlow, 1982). However, textural features play an important role in face recognition and probably represent an important coding dimension in face perception (O'Toole, Vetter, & Blanz, 1999; Russell, Sinha, Biederman, & Nederhouser, 2006), and our results are consistent with similar norm-based code for both textural and configural dimensions.

The age aftereffects we found are also notable for implying a high-level and possibly face-specific site for the adaptation. Analyses of our images confirmed that simple textural properties including mean color and contrast did not vary systematically with age. But more importantly, such features could not be the basis for the aftereffects, because they would not predict opposite aftereffects for old and young adaptors, or no adaptation for faces judged as a neutral middle age. Whatever the stimulus cues driving the effects, that the norm for the adaptation parallels the perceptual norm for age implies that the aftereffects reflect changes in the actual visual representation of age (Webster & Leonard, 2008).

# II. Sibling Rivalry

**Abstract**

We used the phenomenon of binocular rivalry as a test for the neural representation of faces. In rivalry, stronger or more salient stimuli dominate. In norm-based accounts of face coding, distinctive faces are stronger because faces are represented by their distance or “contrast” relative to an average prototype, which itself induces a minimal response. In exemplar-based accounts, each face is coded in terms of its absolute values along a set of dimensions, and thus average and distinctive faces should generate similar activity levels. To compare these, we pitted a neutral face against a distorted version of the same face, which varied the distinctiveness while keeping low-level image features comparable. Observers viewed the images through a stereoscope and monitored the rate of switching between the eyes and the duration of different percepts. Distorted faces were perceived significantly more often, and we show that this dominance cannot be accounted for by an alternative “atypicality” explanation for the effects, in which the two faces fuse to form a blended average which is itself judged more like the distorted face. These results are consistent with norm-based models of facial distinctiveness.

**Introduction**

How the visual system represents faces remains unresolved. The two primary models that have been proposed are prototype-referenced (norm-based) codes and template or exemplar-based codes (Valentine, 1991). These primarily differ in whether the face is represented using a relative code, in which each face is represented by how it differs from an average face  (the prototype or norm); or an absolute code, in which the characteristics defining the individual face lead to a local peak in the distribution of responses encoding different dimensions of the face. An important difference between these models is how facial distinctiveness is encoded. Distinctiveness can correspond to properties such as a more intense expression, or a more unusual appearance.

A hallmark of norm-based models is that faces which are more distinctive (i.e. less like an average face) are represented as further from the norm or center of face space. Because these differences are represented as an intensive dimension or "contrast", they correspond to higher levels of activity in the neural representation.  Exemplar-based codes instead represent distinctiveness by the properties of the population responses (e.g. the sparseness of the sampling), but again code each face by which channels are responding (rather than how much). For exemplar-based models there is nothing intrinsically special or neutral about the average face.

Several recent lines of evidence have been taken to support norm-based coding of face. These include caricature effects, in which the identity of a face is exaggerated specifically by amplifying how it differs from the average (Jeffery & Rhodes, 2011). Distinctive faces have also been found to induce stronger BOLD responses in the fusiform face area (Loffler et al., 2005). Moreover, single cell recordings of face selective cells tend to give monotonic responses to facial distinctiveness with the weakest response to faces near the prototype (Leopold et al., 2006). Finally, studies of face adaptation have largely been taken to confirm norm-based representations (Webster &

MacLeod, 2011), though some aftereffects have also been used to argue for more distributed, exemplar-based representations.

In the current study we explored a novel test of norm-based vs. exemplar-based models of face perception, by using the phenomenon of binocular rivalry (Breese, 1899, 1909). When sufficiently different images are presented to the left and right eyes, the brain cannot fuse them into a single percept, and the observer instead experiences a random alternation between the two stimuli or a patchwork drawn from the two images. The dynamics and characteristics of rivalry have been extensively studied (Blake & Wilson, 2011). Importantly, the switching between the eyes is not completely random, for often one stimulus (or one eye) can be "dominant." For example, when one eye views an image with higher luminance or chromatic contrast, that image is typically seen more often (Knapen, Kanai, Brascamp, van Boxtel, & van Ee, 2007; Mueller & Blake, 1989). Other examples are that rapidly flashed objects capture dominance quickly (Wolfe, 1984), and even more so if observers are aware of the presented image or are adapted to it (Hancock, Whitney, & Andrews, 2008). A further important finding is that dominance can depend on both low and high-level properties of the images, suggesting that the dynamics of rivalry reflect processing along many stage of the visual stream. For example, more recognizable stimuli are more dominant, and in particular faces are dominant over objects or noise (Yu & Blake, 1992). The rate of rivalry is also dependent on perceived aspects of the image, and not the actual physical stimulus.

With faces, a number of studies have shown that dominance depends on expression strength. Fearful and happy faces dominate over neutral expressions (Bannerman, Milders, De Gelder, & Sahraie, 2008), and positive expressions tend to be seen more than negative expressions (Yoon, Hong, Joormann, & Kang, 2009). Expressive faces are also more likely to be visible under continuous flash suppression (Gray, Adams, Hedger, Newton, & Garner, 2013), a technique in which a rapidly changing montage of squares is shown to one eye in order to suppress the visibility patterns presented to the other eye.

In the present study we tested for analogous dominance effects for facial identity, by pitting a normal face against a distorted version of the same face (Webster & MacLin, 1999). An advantage of varying distinctiveness in this way is that it allowed us to create faces that clearly differed in how distinctive or atypical they were, while only minimally changing the low level characteristics of the images. Our results suggest that distorted faces do significantly dominate in rivalry, and we argue that this dominance is consistent with a norm-based account of the encoding of the facial distortions

## Method

### *Observers*

Observers were undergraduate and graduate students at the University of Nevada, Reno. Participation was with informed consent and all testing followed protocols approved by the university's Institutional Review Board. All observers had normal or corrected-to-normal binocular vision and no color difficulties as assessed by standard screening tests. Ten observers were used in Experiment 1, 11 in Experiment 2, 12 in Experiment 3, and 10 in Experiment 4. Some observers participated in multiple experiments.

*Stimuli and Equipment*

Frontal-view images of a male and female faces with neutral expressions were selected from the Radboud Faces Database  (Langner et al., 2010). These faces were then distorted by locally contracting or expanding the features following procedures described in Webster and MacLin (Webster & MacLin, 1999) (Figure 4). Each face used in this experiment was distorted by compressing or stretching vertically or horizontally, creating "sibling" pairs. The same ten faces were used in all experiments.

**Figure 4: Example of the stimuli used in the Sibling Rivalry experiment. One image is shown to each eye, and observers indicated what their current percept was each time it shifted.**

In Experiments 1, 3, and 4, the experiment was controlled by an Apple Powerbook running MATLAB and Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Video output from the laptop went to a SONY Trinitron 500PS CRT monitor rendering at 86 Hz. Observers' view of the CRT was separated by a septum and mirror stereoscope that divided the screen vertically into left and right halves and limited each eye's field of view. Observers viewed the stimuli from a distance of ~155cm, seated and using a chinrest. In three experiments, each stimulus combination was presented in pseudorandom order an equal number of times to each eye. Observer responses were recorded using the numbers on a standard keyboard.

In Experiment 2 stimuli were displayed on a SONY Trinitron E540 viewed non-dichoptically from 114 cm away. Responses were recorded using an external numeric keypad.

*General Procedure*

In the rivalry experiments, normal and distorted versions of the same face covering approximately 3.7° x 4.4° of visual angle were presented dichoptically creating a stimulus that can lead to binocular rivalry (Figure 4). Observers viewed each pair for at least 60 sec, and the trial terminated with the first button press that occurred after 60 seconds had elapsed. During the trial the images remained static on the screen. Nevertheless, their percepts varied over time, and these variations were indicated by key press every time their percept shifted and what the percept shifted to. This allowed measures of dominance for each physical stimulus, which was represented using three possible measures. First, the **total** time spent in a state was measured. If the observer started seeing one percept and switched halfway through the 60 second trial, we would record the total time as 30 seconds for each percept, indicating equal strength of dominance. We also recorded the **mean** time spent in a state. Here a single uninterrupted 30 second period of seeing one face would have a longer mean time in that state than 30 blocks that are exactly 1 second each, although both would have identical total times. Finally, we could **count** the number of times the percept switched, irrespective of duration. These measures provided overlapping information about which percept is dominant, and whether this dominance corresponded to seen the most, seen the longest, or seen the longest average derived time. Each observer repeated the experiment 4 times for each pair of faces and their distortions, so that observers ran 40 trials total. The eye/face combinations were counterbalanced across trials to remove effects of eye dominance. Results were averaged across all face distortions and both face sets.

*Procedures specific to experiments*

*Experiment 1.* The goal of Experiment 1 was to determine whether a distorted face is more dominant in a condition of binocular rivalry. Observers pressed buttons on a keyboard to indicate which of three percepts they were currently seeing: the normal face, the distorted face, or a percept that corresponds to neither face. This last response may correspond to loss of binocular fusion or to another percept such as a piecemeal combination of the two faces. Observers were instructed to indicate their current perceptual state using a single button press immediately upon shifting to a new state.

*Experiment 2.* This experiment tested the category boundary for normal vs. distorted judgments. Observers were presented with versions of the same 10 face exemplars used in Experiment 1, but they were viewed binocularly in a non-rivalrous condition without the use of a stereoscope. On each trial, three faces each subtending 5.8° x 6.8° were horizontally displayed. The left and right faces were non-distorted and fully distorted versions of the same face, and were selected from one of the 10 exemplars used in Experiment 1. The center face was randomly selected from a pool of seven partially distorted versions of the same face (range: 20%-80% distortion, steps in 10% increments). Distortions were both vertical and horizontal and identical to the other experiments. Observers indicated using key press whether the partially distorted face more closely resembled either the neutral face or the fully distorted face. Faces were displayed until a response was made. Each distortion level of the center face was tested 24 times for each of the ten face exemplars. These exemplar x distortion level combinations were presented in a pseudorandom order for a total of 1680 trials. A sigmoid function was fit to the results to estimate the distortion level at which the observer was equally likely to make either response (i.e. probability of 50%).

*Experiment 3.* Rivalry effects were retested while allowing observers a greater range of alternatives to describe their responses. Two new choices were added to the previous ones, with examples of the stimuli corresponding to each shown prior to testing. The new options correspond to the pixel-average of the two face-images (e.g. a "low-level" averaging) or the average morph or distortion level between the two faces (e.g. a high-level" averaging) (Figure 5). A training period was shown prior to each trial in order to help observers learn the differences between the expected stimulus responses. During training, observers were first shown examples of the four potential percepts for that face and prompted for the appropriate response. Then they were shown the four percepts in random order and had to correctly classify them to proceed. After two training periods prior to each trial, the experiment proceeded with 60 s of binocular rivalry over 40 trials total, 4 for each exemplar. All faces in this experiment subtended 6.8° x 8.5°.

**Figure 5: Possible responses in Sibling Rivalry Experiment 3. The top two images are the neutral and fully distorted faces, and were the only physical stimuli actually presented in the stereoscope. Observers had the option of indicating that they saw one of these extremes, or that they saw the average based on high level face space properties (lower left) or low level pixel average (lower right).**

*Experiment 4* was designed to limit fusion of the two faces, and thus minimized similarity of features which may increase fusion. Two manipulations were done to limit fusion. First, the faces were rotated in opposite directions, ±20° from vertical, or 40° apart. Then the faces were cropped with an elliptical mask to limit fusion of external features such as the hair (Figure 6). These features might provide an anchor point as they are similar if not identical in both faces, existing in a region of the image where the Gaussian taper of the distortion reduced any changes between the two stimuli to negligible levels. Responses in this task were changed so that observers now indicated whether they were currently seeing a right-tilted (clockwise) or left-tilted (counterclockwise) face. Their responses were later correlated to perception of a distorted or neutral face.

**Figure 6: Stimuli used in Experiment 4. Each is tilted 20° from vertical, and hair is removed or minimized to limit fusion of the two faces. Observers would judge the non-distorted image on the left as right or clockwise tilted and the distorted image on the right as left or counterclockwise tilted.**

## Results

In Experiment 1, distorted faces were found to be significantly more dominant in rivalry compared to neutral faces using a one-tailed t test (Figure 7). In this and subsequent experiments, refer to figures for full significant differences. Relationships marked ns or unmarked This significance was true for all three measures of: total time ($t(9) = 3.58$, $p < 0.01$), mean time ($t(9) = 2.22$, $p < 0.05$), and the count number of times observers switched to a given state (t(9) =) 2.86, $p < 0.01$. The “neither” judgment was also pressed frequently, suggesting perception of something other than the two stimuli physically presented.

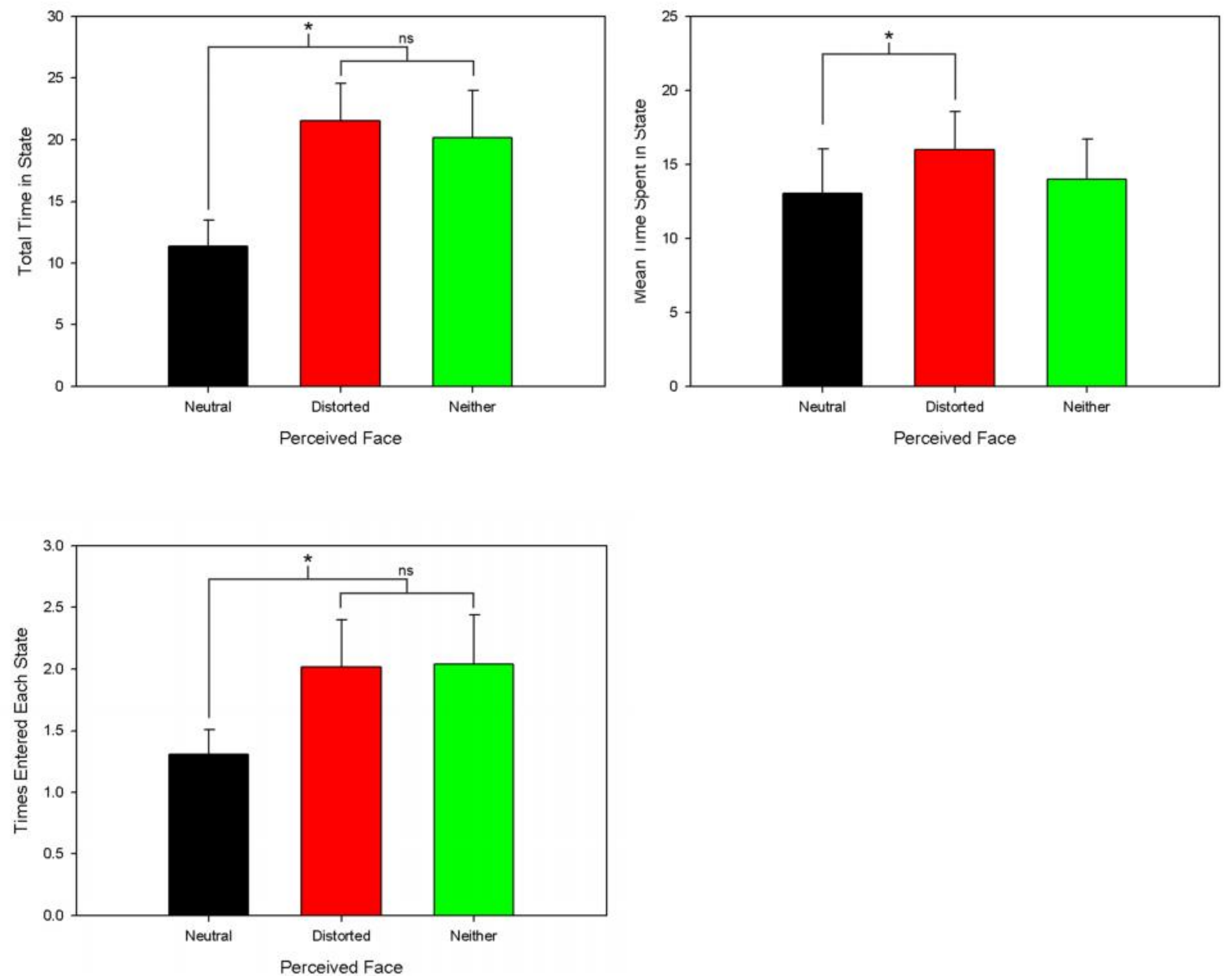


**Figure 7: Experiment 1. Compared to normal faces, distorted faces were seen for longer overall (top left), when distorted was seen observers stayed in that state longer (top right), and they switched to that state more often (bottom left).**

Experiment 2 was designed to assess the influence of "atypicality" effects on the observers' responses (Tanaka & Corneille, 2007; Tanaka, Giles, Kremen, & Simon, 1998). In these, a face half-way between a norm and distinctive face often appears like the distinctive face. This bias could provide an alternative account of the dominance found in Experiment 1. Specifically, suppose that observers were simply fusing the images on a number of trials, and thus perceiving this fused stimulus. The bias toward calling this average more like the distorted image could then simply reflect the bias in the judgments, and not in the degree of dominance. The observed judgments confirmed the atypicality effect for our conditions, but suggested that it was relatively weak. The raw data represented the percentage of trials where a face was judged non-distorted, and a probit analysis was used to find the point of subjective equality, where observers would have a 50% chance of designating a face as either "normal" or "distorted." This analysis created a fit to the linear equation y = 0.0833x - 3.6667. In the absence of any bias, this would fall on the 50% morph level between fully normal and fully distorted. Instead it was found to be at about 44% distortion (Figure 8). This means that a wider range of distortion levels were classified as distorted than those that were classified as normal. Again this suggests a bias towards seeing the distorted face.

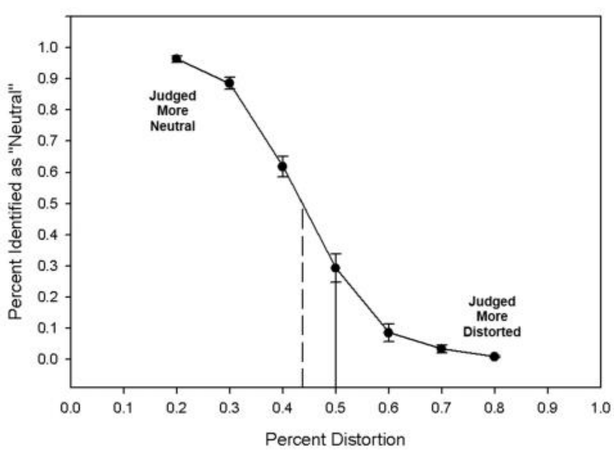


**Figure 8: Experiment 2. Classifications of partially distorted faces as being closer to either the "normal" or "distorted" percepts. If there were no bias towards perceiving a distorted face over a normal face, the 50% distorted face would have an equal chance of being seen in either category (solid line). Instead, the point of subjective equality was at a face that would be approximately 44% distorted (dashed line).**

Experiment 3 introduced two more response button choices to further investigate the possibility of stimulus fusion. With few exceptions, the responses were split into two significantly different categories: normal, distorted, and middle distort vs. pixel average and none (Figure 9). Within these groups, the first comparison of interest was normal vs. distorted, and there were no significant differences among any of the three measures. Therefore, there is no longer a bias toward reporting the distorted image more often as in Experiment 1 (perhaps because Experiment 1 did not provide a sufficient range of responses to fully characterize the observers' percepts). Also of interest were the relative frequency of the two new choices; both were reported by the observers, but the average distortion was reported far more frequently than the pixel average.

Experiment 4 used techniques to limit fusion of the two images and found that the distorted vs. non-distorted bias partially persisted. It remained for the total measure, $t(9) = 2.89$, $p = 0.02$, but was not significant for the count number, with all percepts seen at roughly the same rate. The mean measure was also not significant, possibly because its value is contingent on the count measure (Figure 10).

## Discussion

Our results show that when normal and distorted face images are pitted against each other in binocular rivalry, distorted faces are perceived significantly more often. One potential account of these effects is that the distorted faces dominate because they have a higher effective "contrast" in the visual representation of faces, and thus are more potent visual stimuli. Again, this is consistent with norm-based coding models of faces, which assume that the neural response to any face increases monotonically according to how much it differs from a neutral or average face. Our results thus may provide a novel test and confirmation of this coding scheme. However, we first need to consider a number of alternative potential explanations of the dominance. The first is whether the measured effects have anything to do with the fact that the images were of faces. As we noted, rivalry depends on the properties of the stimuli at many points along the visual stream, beginning at early levels where the relevant determinants of dominance depend on simple properties such as the overall luminance and color and contrast in the images. However, our stimuli provide a built-in control for these image factors. An advantage of using distortions is that they allowed us to isolate and systematically vary the distinctiveness of the faces without changing low-level features. The distortions themselves are not equivalent to a change in face identity, but parallel it in face space. Moreover, adaptation studies have shown that adapting to distorted faces transfers across changes in the position, size, or orientation of the faces, and thus at least partly arises at non-retinocentric levels of the visual representation (e.g. (Watson & Clifford, 2003; Zhao & Chubb, 2001)). Finally, as detailed in the Introduction, several studies have revealed that rivalry and dominance can arise at face-specific levels, such that dominance can be directly related to differences in expression strength. Collectively these points make it plausible to assume that the effects we observed may be driven by the perceived characteristics of the images as faces, and that the therefore depend on how the visual system encodes and represents the images as faces.

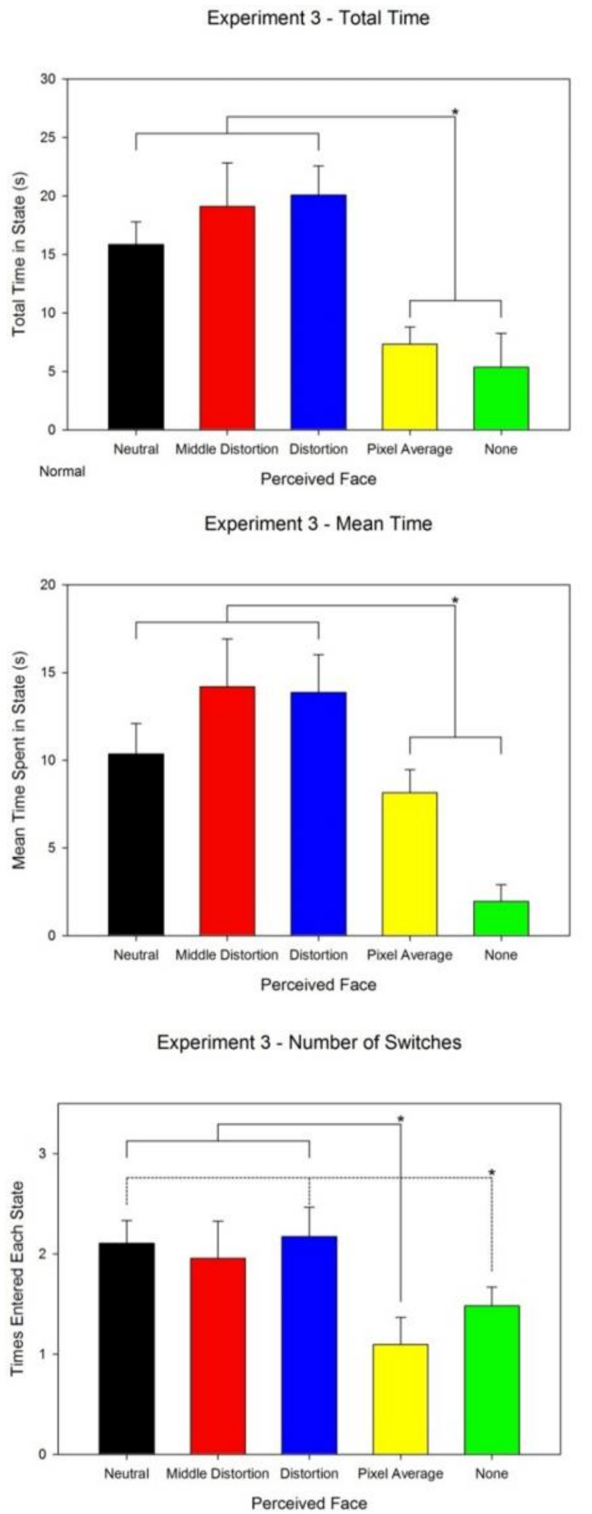


**Figure 9: Experiment 3. Percept settings when observers are given more response choices. Distorted dominance bias has disappeared, and the middle distortion percept is often seen.**

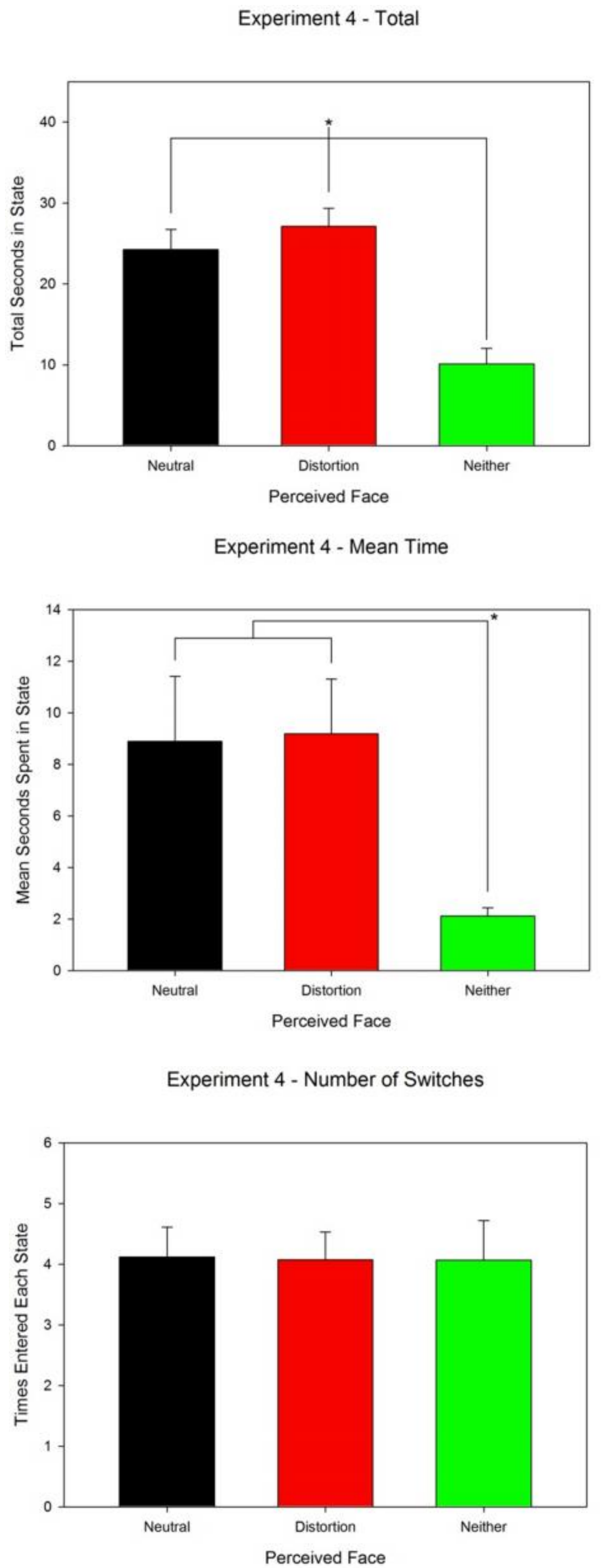


**Figure 10: Experiment 4. Significant differences between distorted and non-distorted were found for the total measure, but not mean or count.**

A second possibility is that the distorted faces did not dominate the rivalrous interaction, but instead appeared dominant because perceptual mixtures of the two faces were more likely to be perceived as distorted. By this account, the stimuli may often have appeared fused rather than rivalrous, but observers were biased in how they rated the fused stimulus. Several lines of evidence are consistent with this. First, when we allowed the observers the option of choosing a blended stimulus to describe their percept, these occurred with the same incidence as either the original or fully distorted face (Experiment 3). Thus it is clear that something like a fused percept was often experienced. It is notable that this blend was at the level of the mean face rather than the simple image average. This could itself suggest a high level interaction, though it might also reflect stereopsis through fusion of the binocular disparities in the display, so that the images were perceived with some depth variation. We cannot exclude this possibility, but note that the fused percepts were reported with similar frequency whether the images were distorted vertically or horizontally. Again, if there were a simple linear averaging of the two eyes' images, then one eye could still appear to dominate if that average was perceived to be more similar to one of the original faces than the other. Experiment 2 was included to test for this possibility, and specifically to test for "atypicality effects" for our stimuli, in which an average of a prototypical and distinctive face appears more like the latter. These effects have also been explained by the properties of face space (Tanaka et al., 1998). By one simple account, they could reflect a saturating nonlinearity in the response to distinctiveness, which also might explain why we are better at discriminating small changes in the faces were are normally exposed to while insensitive to large changes in faces far removed from the norm, the "other-race" effect (Tanaka, Kiefer, & Bukach, 2004). Our measurements confirmed a weak "atypicality" bias for our distorted faces. Thus we obtained evidence for both frequent fusion and a bias toward the distorted face when fusion occurs. However, the one argument against this account is that the biases persisted even when we introduced sizable orientation differences between the two faces in order to minimize fusion (Experiment 4). The fact that the dominance of the distorted face survived this manipulation suggests that fusion and atypicality biases could not provide a full account of the rivalry effects, and that they instead depend at least in part on the inherently higher visual salience of the more distinctive face.

In conclusion, the present results are consistent with rivalrous interactions between faces that depend in part on the relative salience or distinctiveness of the faces, which depends in turn on the norm-based coding scheme that the visual system uses to represent faces. In color vision, colors that are farther from the norm (gray) are more saturated, and color opponent mechanisms code more saturated colors by larger deviations from their null response. Essentially the same principles have been used to explain how more "saturated" faces are represented in a norm-based code (Webster and MacLeod, 2011). We have shown that the pattern of binocular rivalry for neutral or distorted faces is also consistent with this principle, thus providing a further piece of evidence for this computational principle.

# III. Perceiving relationships within face space

**Abstract**

Most theoretical models of face perception assume that faces are encoded within a "face-space" that represents each individual by how they differ from a prototype. These models can account for many aspects of face perception and how it varies with metrical changes in the stimulus (e.g. with changes in angle or distance from the center). We asked whether observers have cognitive access to this representational framework, by asking whether they can correctly perceive the relationships between different vectors within face space. To limit the dimensions, the empirical space was restricted to two-dimensional spaces corresponding to pairs of expressions and their anti-expressions (faces that were on the opposite side of a neutral expression relative to one of the basic expressions). In one task, observers were asked to adjust the angle defining a target face so that it was "opposite to" a reference face. These settings were highly variable, showing that observers in fact have little intuition about how one expression is related to another. In a second task, observers compared the magnitude of different expressions. They turned out to be good at making these cross-category intensity judgments for actual expressions, but not for their anti-expressions. Finally, in the third task, observers selected the average of a pair of faces that differed in their angle within the space by 90°, but had the same expression intensity. In this case most observers could reliably select an average, and notably this average corresponded to the Euclidean mean of the pair, suggesting that they were not representing angle and magnitude independently.

Together these results suggest that while a space-like coordinate system may underlie the visual representation of faces, observers do not have conscious access to this system in a way that could allow them to judge the metrical relationships between different faces.

## Introduction

Standard models of face perception hold that we represent faces with a relativistic code in which each face is represented by its direction and distance from an average prototype. This model captures a surprisingly wide range of evidence about face recognition and discrimination. For example, it explains why caricatures correspond to faces that are amplified in magnitude from the mean, but not when the face is changed in direction (which instead appears as a new identity) (Lee et al., 2000). It can also account for a large body of results on adaptation effects in face perception, which behave as though the space is renormalized relative to the current stimulus. (Webster & MacLeod, 2011) Finally, it is also consistent with the monotonic response functions with increasing distinctiveness in face selective neurons observed with fMRI or single unit recordings (Leopold et al., 2006; Loffler et al., 2005).

This coding scheme for faces is not unique and instead reflects a very general strategy in perception. For example, the variations in face identity or identity strength with changes in angle or magnitude relative to a prototype is remarkably similar to standard models of color vision, in which color opponency represents any color by its angle (hue) and distance (saturation) relative to an average gray (Webster & MacLeod, 2011). The notion of such perceptual spaces is thus widespread in perception.

In this study, we asked whether observers have “cognitive access” to this framework, such that they have intuition about the vector relationships between different coordinates in the space. Our work was prompted by a recent study by Webster et al. showing that individuals are remarkably poor at judging which pairs of colors are complementary (opposites) (Webster, Nolan, Sternberg, Kay, & Webster, 2013). Thus even though these opposite colors are a hallmark of color opponent theory – and are in theory encoded by equal and opposite neural responses – this relationship is not consciously available to an observer. Here we asked whether similar behavior would be observed in face space, where like color complements, faces and their anti-faces have come to play a special theoretical role in models of face perception (Leopold 2001; Rhode and Jeffrey, Vision Res 2006?).

A clear difference between color and face spaces is that the latter must be much higher-dimensional, and these dimensions remain largely unknown. To reduce this problem, we instead focused on the space of facial expressions. Since Darwin (1872), expressions have been thought to be largely innate, and may reflect only a few principle or basic dimensions. Ekman and Friesen (1971) provided evidence for a universal set of basic expressions which Ekman eventually narrowed it down to six: happiness, sadness, anger, fear, surprise, and disgust. Each of these is characterized by specific behavioral responses, for example a person who encounters something disgusting will recoil from it, lower their brow, and sneer.

Later work led to many more cross cultural studies (Ekman et al., 1987, 1969; McAndrew, 1986). Providing further evidence that expressions are processed as discrete categories, and faces are judged by their proximity to each expression, perhaps as a norm-based strategy that selects the "best fit" (Leppänen, Richmond, Vogel-Farley, Moulson, & Nelson, 2009; Young et al., 1997). It may be that some face processing is developed over time through experience, but the willingness to attend to and classify faces appears innate. For example, it appears that infants preferentially attend to faces, although their understanding of fearful faces doesn't seem to develop until about 7 months (Peltola, Leppänen, Mäki, & Hietanen, 2009). Additionally, developing an understanding of the *differences* between emotions develops gradually, but the expression face space structure exists from an early age (Gao, Maurer, & Nishimura, 2010).

Adaptation studies have strongly implicated a norm-based code for facial expressions (Skinner & Benton, 2012). Adapting to an expression causes judgments of a normal face to seem less like the adapting face (it moves in the opposite face space direction), and there is minimal transfer across expression (Juricevic & Webster, 2012). This suggests that the different expressions are represented independently (or as orthogonal axes) in face space. Furthermore, expression aftereffects partially transfer when adapt and test identity are incongruous, giving credence to expression as an independent face category (Fox & Barton, 2007; Skinner & Benton, 2012).

A second reason for exploring expression space is that – again like color – what constitutes an opposite stimulus may change from one level of the visual system to the next. The color opponent mechanisms in the retina and geniculate are not the “red vs. green” and “blue vs. yellow” dimensions that characterize Hering’s opponent channels. Analogously, one could define opposite expressions by different criteria. Specifically, if one were asked to choose the opposite of a “happy” face, one possible response might be to name one which is conceptually opposed from an emotional standpoint, typically “sad” or “angry.” Alternatively, anti-expressions are typically assumed to reflect opposite facial configurations. For example, many analyses have been done to determine the exact facial muscle movements needed to create a specific expression (Ekman, Levenson, & Friesen, 1983; Levenson, Ekman, & Friesen, 1990). If these movements could be reversed, you would instead have the shape opposite of the expression; an anti-face in reference to a specific expression, or anti-expression. Thus it is of interest to ask not only whether an observer can actually estimate and opposite, but what space they might be working in.

In a second set of experiments, we also examine the perception of the relationships between differences in vector magnitude or stimulus intensity. For emotional expressions, this is again analogous to color saturation or distance from the norm, and here observers are surprisingly good at judging the relative contrasts along different dimensions of color space (Switkes & Crognale, 1999). We used similar tests to explore whether we can recognize the relative intensity of faces and anti-faces.

A final experiment considers our ability to make across different angles and distances, by asking observers to estimate the average of two faces. This perception of the mean has been recognized as a fundamental encoding process in both low-level and high-level domains, including the perception of faces (Haberman, Harp, & Whitney, 2009; Whitney, Haberman, & Sweeny, 2013). Here we specifically ask whether this averaging for expressions reflects Euclidean averaging or whether the dimensions of distance and angle might instead be averaged independently.

## Experiment 1: Complementary Faces

## Method

*Observers*

Observer participation in all these experiments adhered to the University of Nevada's Institutional Review Board standards. Eight observers made angry-happy matches, 5 fear-angry, 7 fear-happy, and 5 happy-disgusted. Some observers did multiple experiments, with 5 who completed all 4.

*Stimuli*

Faces were generated using the computer program FaceGen 3.3 (Singular Inversions). This software allows the user to generate high-resolution images of human faces. Specific attributes can be adjusted, such as gender, race, age, and expression type, and these can be set to specific levels, such as 75% male or 40 years old. FaceGen also has sliders for various facial manipulations, among which are 7 sliders approximating the 6 basic expressions (these experiments used the *SmileOpen* slider for an open mouth happy, but there is also a *SmileClosed* option which is subtler). Although some sliders are not independent (e.g. race), the expressions can all be adjusted together or separately, so that one could make a face that is both angry and fearful, or even some unequal proportion of the two.

Many of the sliders can be manipulated by typing in specific values (0.00-1.00, steps of 0.01). Lowering the level below 1.00 causes the specific expression to decrease in intensity in a realistic manner. FaceGen also allows another type of face to be created: anti-faces. These faces represent the face space opposite of a given face, linearly passing through the normal, unemotional face exemplar in face space (Leopold et al., 2001). If you were to quantify the manipulations to change a fearful face into a neutral face, and then continued to apply them past the neutral face, then an anti-face would be created. In FaceGen, sliders can easily be given negative numbers, so -0.50 is the anti-face in reference to +0.50 in a given expression category. The anti-face in this experiment represents selection of a pure shape opposite.

To make the images, first an average face was created using the “Set Average” button on a face set to “Male,” as opposed to “Very Male,” roughly 44.4% on FaceGen’s male to female scale. The age was set to 30, the race was mixed/equal levels, and caricature and asymmetry were minimal. Face pairs were created between several possible pairings – angry-happy, angry-fear, fear-angry, and happy-disgust. All were set to 50% (0.50). Anti-faces were also created, so that each individual trial set has 4 extreme faces on two axis continua, e.g. angry vs. anti-angry and happy vs. anti-happy at 0°-90°-180°-270° steps (Figure 11). Next the intermediate angles were created by using the cosine and sine functions to obtain the x and y coordinates which were mapped to the two axes selected for that trial. Therefore, in an angry/happy condition, 45° would be equal amounts of angry and happy, or 0.35, 0.35 in the space @ 50% max expression; 225° would be equal amounts of anti-angry and anti-happy). This design implies that the relationship between two expressions is orthogonal.

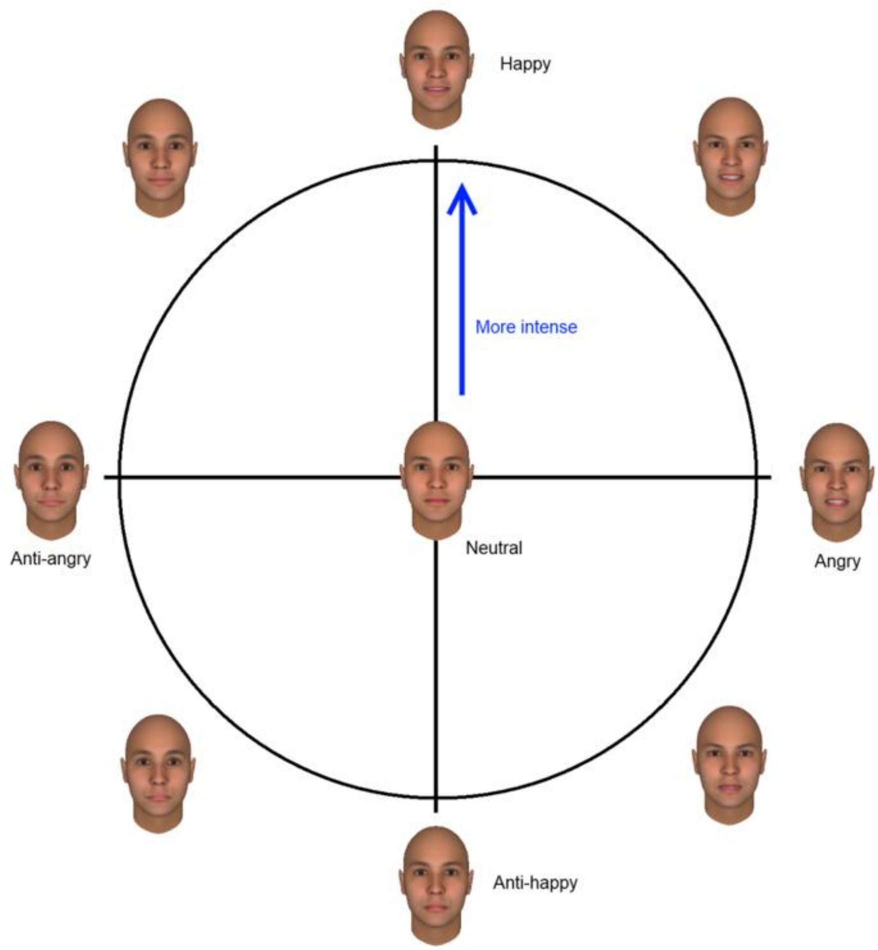


**Figure 11: Face space representation of angry, happy, and their respective anti-faces. Shown are faces at 45° intervals, actual experiment uses less coarse steps. Other paired relationships are represented in a similar fashion, with the first named face at 0°.**

*Procedure*

Observers were presented with two side by side faces, both at a visual angle size of 3.48° x 5.52°. The faces were displayed simultaneously for 500 ms, separated by an ISI of 1500 ms. The reference and test face both appeared in either position half of the time. The reference face was chosen prior to each trial and remained constant within the trial. All face intensities were set at a fixed 0.50 distance from the normal face, and the angle was selected randomly from 8 discrete locations 45° apart in the specific slice of face space (e.g. 0° = happy, 45° = happy/angry, 90° = angry, 135° = angry/anti-happy, 180° = anti-happy, etc.). The experiment was repeated 10 times for (8 * 10) 80 trials.

Observers were told to adjust that face's appearance along the same circular face space continuum until they reached a face that appeared to be the opposite of the reference face (in their estimation) and select it (method of adjustment). They were instructed to find "the opposite," but were not given any instructions on whether to find the face space opposite based upon its shape (face/anti-face), conceptual links (happy as the opposite of sad), or given any other leading instructions. Regardless of the criterion used to make a choice, observers made 10 settings for each reference point, in part to see if they could create a reliable pattern of responses (with low variability). To avoid biasing judgments, observers were not given names to the expressions or to any particular face relationships, and the expressions used in each pair were not specifically named, although some observers might have recognized the pairs used.

*Analysis*

Multiple angles were averaged by using the mean of circular quantities due to the data being cyclical. Each angle setting was converted to Cartesian coordinates. The mean of these coordinates was then converted back to angle using the arctangent function. Standard deviations and standard errors were calculated using the deviations of each point from the corresponding mean. These descriptive statistics were calculated for all 8 reference angles and all observers. The settings were then plotted against a line of 180° which would correspond to (shape-based) settings.

**Results**

Figure 12 shows settings made by two observers on the same judgment. If an observer had selected the complementary face at each angle, representing complete shape-based judgment of the opposite, then these lines would all pass through the center (as shown for the theoretical matches at 45°-225° and 135°-315°, which reflect perfect opposites). As illustrated, this is sometimes found in the empirical results but not with any specificity or regularity. Instead, the choices appear to vary somewhat chaotically and often reflect angles much closer than the actual complements.

The variability (standard error) was used as a measure to determine observers' accuracy in each trial. One way to measure this is within observers. Averages were made of all the standard errors derived from 10 trials. They were: 14.69° for angry-happy, 15.59° for fear-angry, 11.35° for fear-happy, and 14.65° for happy-disgust. These numbers are rather high, especially considering it means that e.g. angry-happy standard *deviation* is 71.68 degrees in face space. Additionally, we can measure the between observers standard error by finding the standard error of all observers' mean settings for each angle, then generating an average of these standard errors irrespective of angle. The between-observers standard error was even higher – means of 25.34° for angry-happy, 47.74° for fear-happy, 26.33° for fear-angry, and 61.48° for happy-disgust. – suggesting that observers were not in agreement with each other on

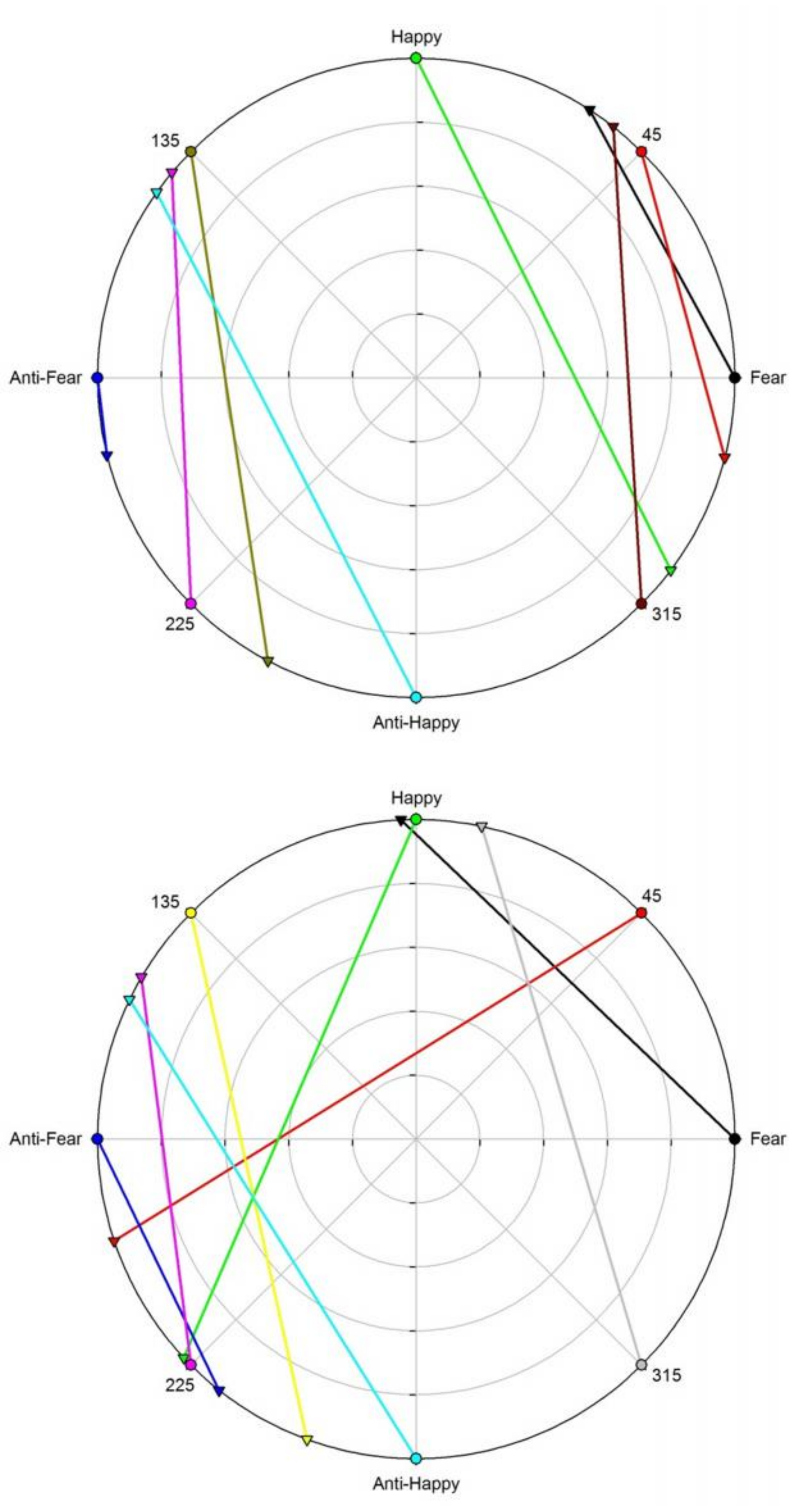


**Figure 12: Example of two observers' settings using reference faces in the fear/happy continuum. Observers were shown one of eight reference faces (circles) and asked to select their opposites (triangles). Settings made ~180° apart would supply evidence of opposition based on shape.**

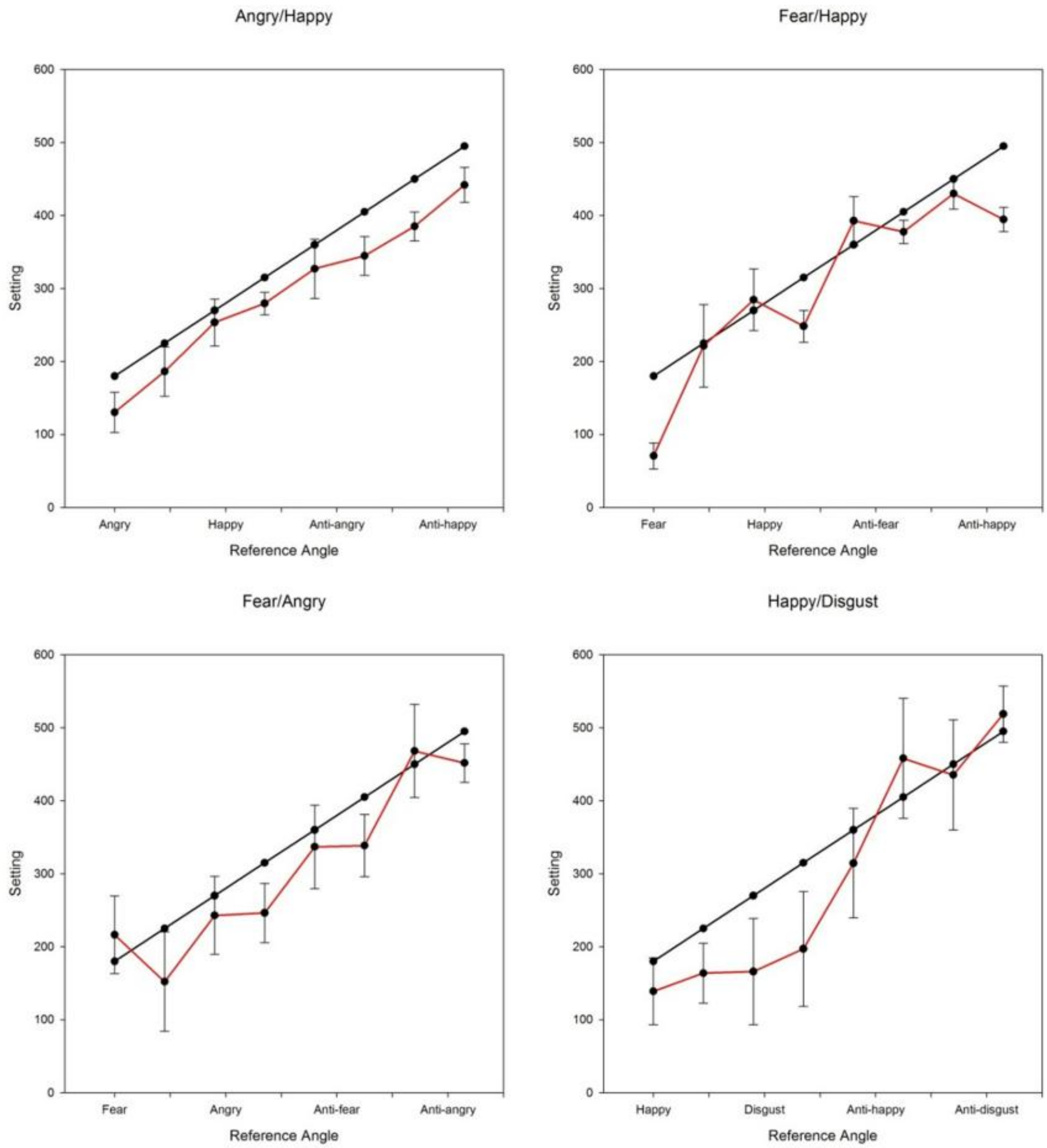


**Figure 13: Opposite face settings (y axis) as a function of reference face (x axis). Intermediate, unlabeled positions indicate a mix of the previous and subsequent expressions. The red line represents average user settings; the closeness of fit to the black line represents how closely observers' judgments fit a 180° match.**

which faces are opposites. These effects are similarly seen when the matches are instead plotted as a function of the reference angle (Figure 13).

Comparable results were found when for space constructed from other pairs of expressions. For example, Figure 14 shows data for four observers viewing fear/happy, and again highlights the differences between observers. In addition, t tests were conducted to see how closely settings fit the line, with a non-significant difference meaning that they did not significantly differ from the configural opposite at that angle. These tests have low power because the observers' variability is so high. Nevertheless, fear-angry and fear-happy settings significantly differed at 6 and 7 out of the 8 angles, respectively, while angry-happy deviated from the 180° only for the last 3 angles. Happy-disgust was not significant at any angle. Comparable analyses for individual observers are listed in Table 1.

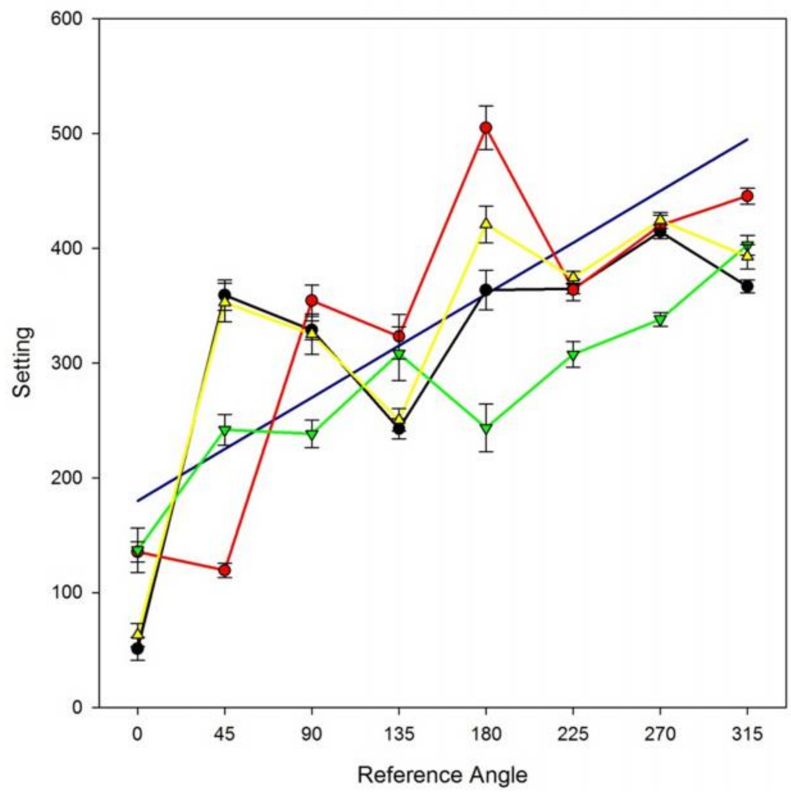


**Figure 14: Dark blue: prediction line for 180° shape opposite settings. Other colors: settings made by 4 observers on a fear/happy pair.**

| | Angry/Happy | | Fear/Angry | | Fear/Happy | | Happy/Disgust | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| **N** | **8** | **%** | **5** | **%** | **7** | **%** | **5** | **%** | **25** | **%** |
| **0** | 5 | 62.50 | 3 | 60 | 7 | 100 | 4 | 80 | **19** | **76** |
| **45** | 4 | 50 | 3 | 60 | 6 | 85.71 | 3 | 60 | **16** | **64** |
| **90** | 3 | 37.50 | 2 | 40 | 7 | 100 | 3 | 60 | **15** | **60** |
| **135** | 6 | 75 | 3 | 60 | 5 | 71.43 | 3 | 60 | **17** | **68** |
| **180** | 7 | 87.50 | 4 | 80 | 5 | 71.43 | 4 | 80 | **20** | **80** |
| **225** | 7 | 87.50 | 5 | 100 | 7 | 100 | 5 | 100 | **24** | **96** |
| **270** | 7 | 87.50 | 5 | 100 | 7 | 100 | 5 | 100 | **24** | **96** |
| **315** | 3 | 37.50 | 3 | 60 | 7 | 100 | 1 | 20 | **14** | **56** |
| **Total** | **42** | **21** | **28** | **14** | **51** | **25.50** | **28** | **14** | **149** | |
| **Mean** | **5.25** | **65.63** | **3.5** | **70** | **6.38** | **91.07** | **3.5** | **70** | **18.63** | **74.5** |

**Table 1: Number of observers that set matches significantly different from 180° opposite predictions ($p < 0.05$) and percentages of observers from the total.**

## Experiment 2: Intensity Matching

### Method

*Observers*

Each condition had 5 observers, of which some were unique and some participated in all conditions.

*Stimuli and Equipment*

101 FaceGen images were created that spanned a range of expression intensity from -0.5 to +0.5 in steps of 0.01. Separate arrays were created for expressions of "happy," "angry," and "fearful." These were chosen as it was determined that they were easier to recognize at lower intensity levels. The faces were displayed in the center of the screen on a black background, with an approximate visual angle of 2.66° x 4.50°.

*Procedure*

A reference face and test face were displayed in succession for 500 ms each, with a 1 s interstimulus interval between each presentation. A beep indicated the test face, which was adjusted in intensity by using a keypad. Observers were instructed to judge the test face relative to the reference, and respond with "stronger than the reference" or "weaker than the reference." They were told to judge each expression's range in reference to itself, e.g. asked if the angry test was more or less angry than the happy reference was happy. This distinction was necessary because at the e.g. 0.15 intensity level, the angry face may have been more apparent than happy. Responses were two-alternative forced choice judgments using a staircase procedure that averaged the last 8 of 10 total reversals. Keyboard input initially changed the intensity by 10 images (~10%

of the image range) and after 2 steps this dropped down to the minimum step of ±1 image (~1%). Two separate task types were completed: between-expression and within-expression. There were 12 between-expression comparisons used: reference/test pairs of angry/fear, fear/happy, and happy/angry; anti-angry/anti-fear, anti-fear/anti-happy, and anti-happy/anti-angry; The other six trials were the same but reversed (e.g. fear reference and angry test). The 6 within-expression trials consisted of angry/anti-angry, fear/anti-fear, happy/anti-happy, and the 3 reversed conditions. All trial conditions used at minimum reference image levels of ±0.1 and ±0.4 (test was adjustable throughout the range), while within-expression trials were additionally tested with references of ±0.2 and ±0.3 in order to make more sensitive comparisons.

**Results**

Averages across all observers are plotted in Figure 15 for expression vs. different expression pairs, and Figure 16 for expression vs. its anti-expression pairs. For cross-expression pairs, observers were able to readily transfer judgments of relative intensity. Reference faces of weak intensity elicited settings that were close to the neutral face, while stronger references caused observers to set strong response faces (e.g. the happier a reference face was, the more fearful a test face was set). The absolute intensity of two expressions were not necessarily matched (thus a +0.40 happy face may have been matched by a +0.30 angry face). This effect scaled with increasing reference intensity. In addition, transference was seen between anti-faces: stronger anti-fear was matched with stronger anti-angry.

When face pairs were tested in intensity transfer within face expressions (e.g. happy reference and anti-happy test, or vice versa), observers had more difficulty making consistent settings. For many conditions, increasing the reference strength did not elicit an increased response. This was true whether the reference was a veridical face or an anti-face. Rather, settings remained at a somewhat constant level.

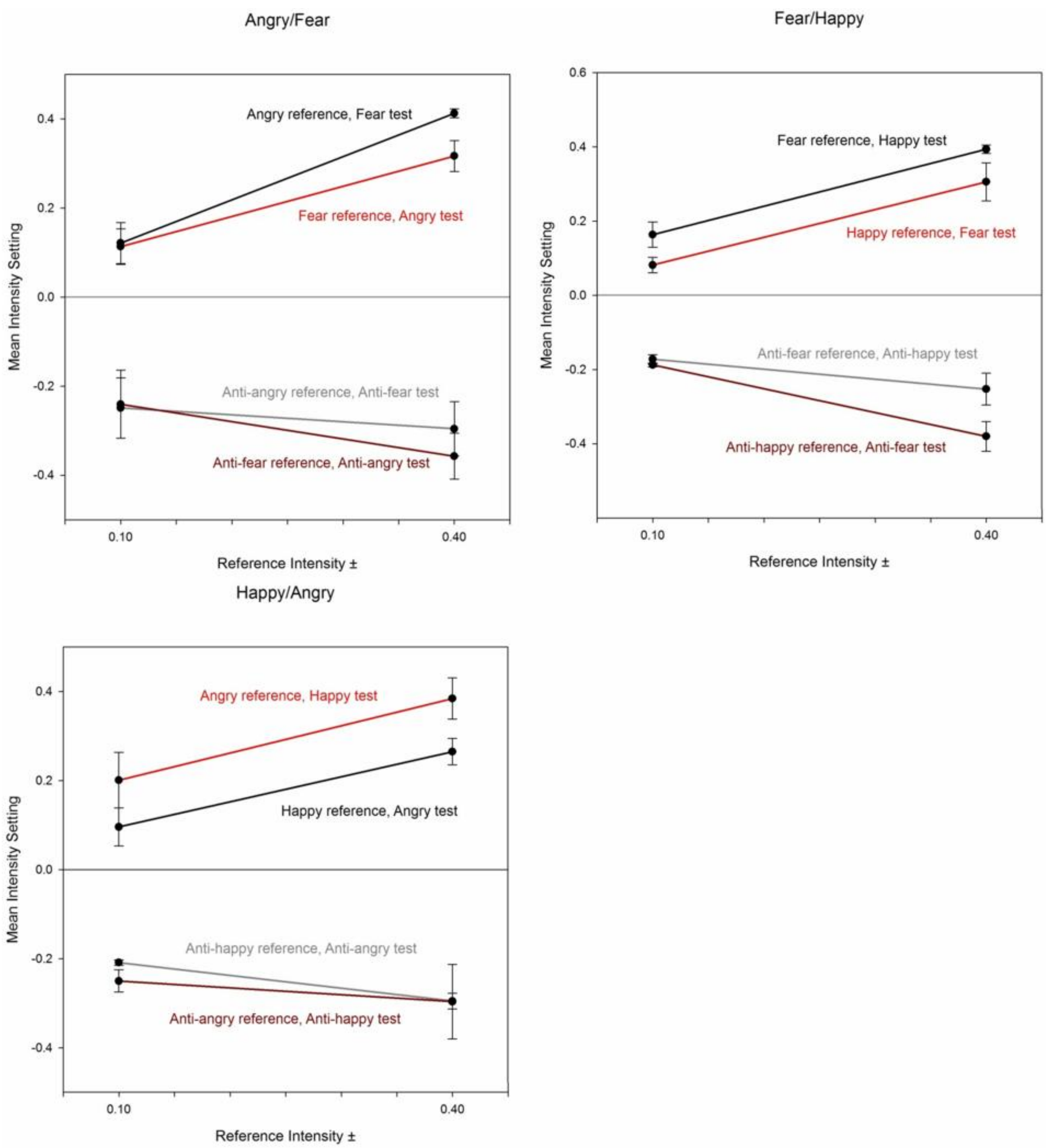


**Figure 15: Averages for between-expression intensity matches. Top left: angry vs. fear, top right: fear vs. happy, bottom left: happy vs. angry.**

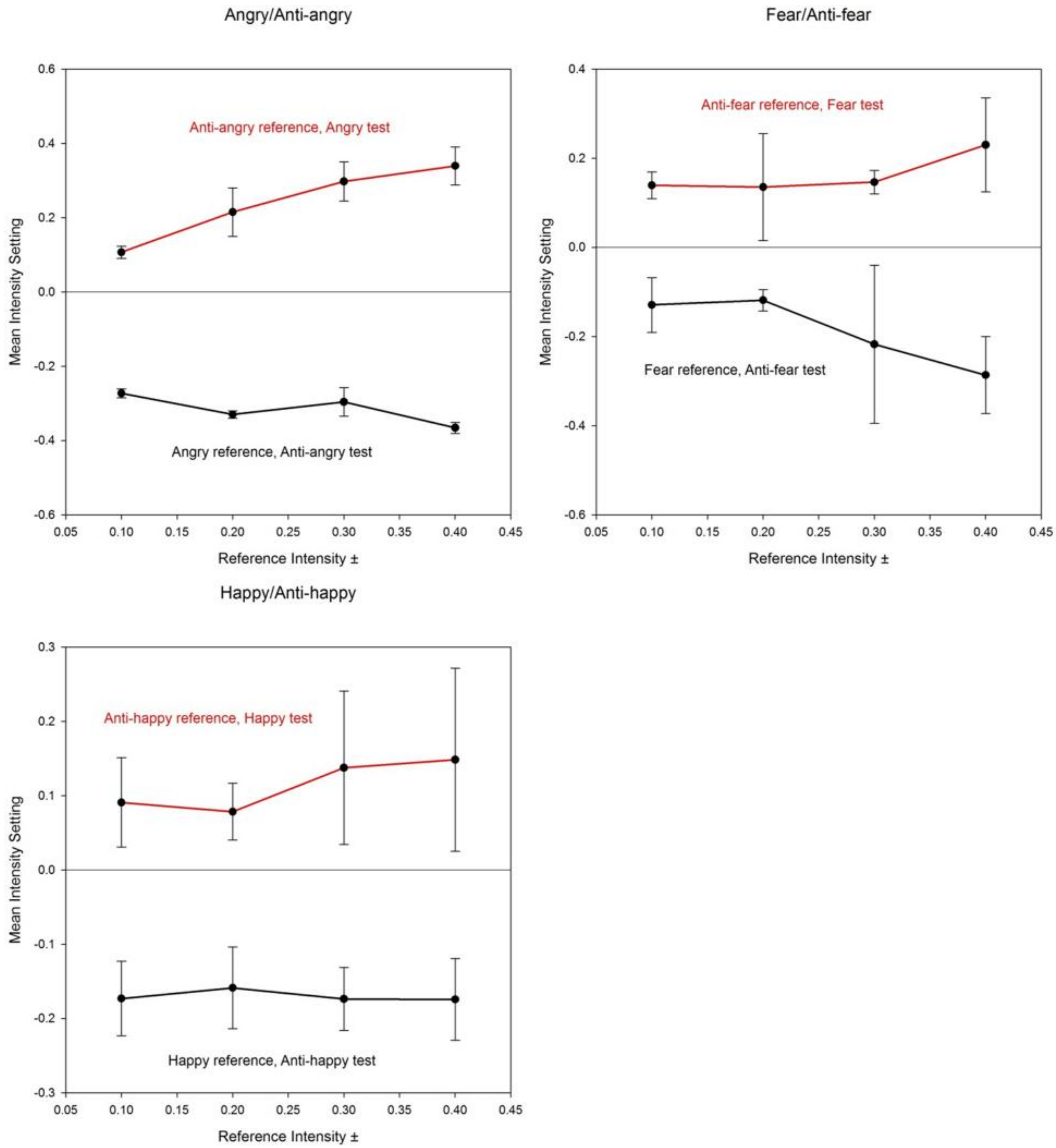


**Figure 16: Settings for within-expression intensity matches. Top left: angry vs. anti-angry, top right: fear vs. anti-fear, bottom left: happy vs. anti-happy.**

## Experiment 3: Averaging Faces

### Method

*Observers*

This experiment used seven observers and identical protocols to previous experiments.

*Stimuli*

An angry vs. happy face set was created in FaceGen as detailed earlier. 401 images were created – intensities from ±0.00 to 0.50 in 0.01 intensity steps at 8 discrete directions in face space, offset every 45°.

*Procedure*

Two reference faces and one test face were used, all 3.48° x 5.52°. The two reference faces were shown side by side for 500 ms followed by a single test face for 500 ms, with 1500 ms ISI following each stimulus presentation. Reference faces were a fixed 90° apart in face space, and for this experiment were fixed at an intensity of ±0.40 (i.e. 80% of the ±0.50 max used for the image range) to avoid ceiling effects and to allow for potential response that are more intense, while preserving a larger range for low intensity settings. The test face was always at an angle midway between the two reference faces, which were ±45° on either side of it (e.g. if the test was 225°, the two references were 180° and 270°). The test face intensity was controlled by the observer. On each trial, the test angle was randomly selected from 1 of 8 directions.

Observers were instructed to look at the two reference faces and imagine an average of them. They were told that the expressions were all different, and so should imagine how strong the test expression would be if it were an average irrespective of the expression. They were to respond as quickly as possible but were given the opportunity to view the faces. Observers used the method of adjustment and indicated when the intensity was perceived to be the average. They had to make at least 5 intensity step adjustments before a decision was allowed to avoid errant key presses or premature selection. After each setting, a new test angle was chosen (with references yoked to that angle) until all 8 settings were made. Observers repeated this process 10 times per angle (80 total settings), which were used to find the averages and standard error. Observers were not made aware of the nature of this face space or the specific intensities or angles that they were shown.

## Results

Mean results are shown in Figure 17, which plots the coordinates of the face selected as the average for each of the 8 different pairs of reference faces. Overall settings were surprisingly linear in face space. For example, when they were shown a 0.40 intensity happy/angry pair at 0° and 90° (coordinates 0.4, 0 and 0, 0.4), the 45° test should have coordinates of (0.28, 0.28) if it was selected to lie along the same circle of constant intensity. Instead, the match intensity was much lower, not significantly different than the (0.20, 0.20) at the midpoint of a linear average.

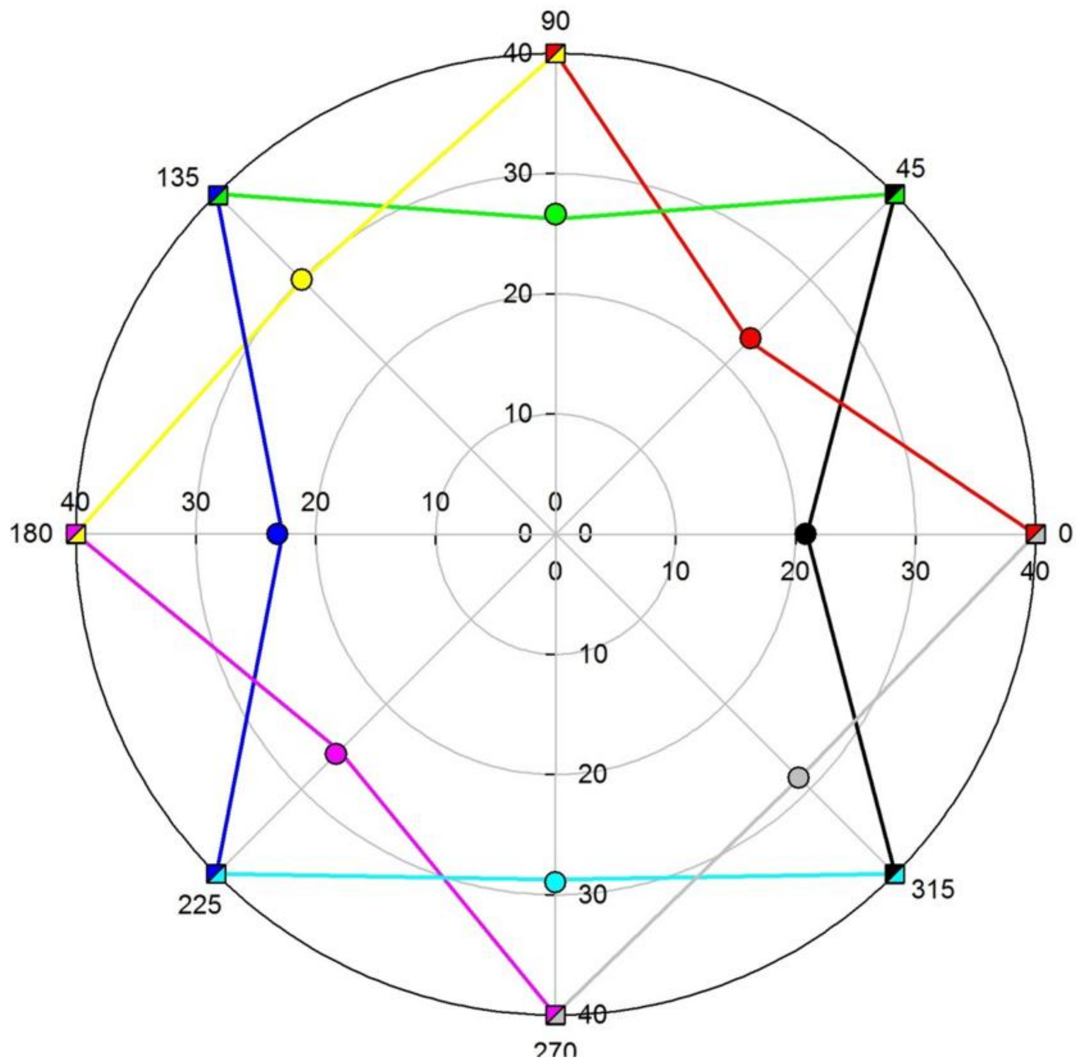


**Figure 17: Intensity settings (circles) as an average of corresponding reference faces presented at fixed intensities (same-color divided squares)**

One observer was an exception and generally made test image settings that were equal to or greater than ±0.40, and thus formed a rough circle in face space. This may be due to a bias towards the ceiling, although this observer was also more experienced with the stimuli (although not initially to the aim of this individual experiment). Three observers made responses strongly indicative of linear averaging, while the other two observers made settings that were lower than the reference intensities, but not indicative of preserving the physical intensity.

## Discussion

Space-like representations are standard metaphor for sensory coding across a wide variety of stimulus domains. In some cases we appear to build explicit "cognitive maps" that encode the coordinates of stimuli in ways that allow us to directly estimate the relationships between them. The obvious case is maps of space itself. We can readily estimate the distance and angles between points, perhaps because the whole point of such maps is to allow such estimates. Other examples are the encoding of orientation or direction, where we are highly efficient at judging the angular relationships between stimuli and thus seem to encode them by their quantitative rather than their qualitative differences. Again the visual system might represent these attributes this way because they remain inherently spatial variables.

But what of perceptual spaces that are not encoding space? Here the perceptual judgments they might afford are less certain. Color is the paradigmatic example of a spatial mapping and represents only three dimensions of the stimulus (e.g. red-green, blue-yellow, or bright-dark). It is well known from multidimensional scaling that people can reliability judge the relative similarity of different colors, but this alone does not require that they literally "see" different colors as different points in a space. In fact, Webster et al. recently found that people are very bad at averaging large color differences (2013). This is not surprising given that it is not intuitively obvious that two opposite colors will average to gray. Yet it perhaps is when considering that this relationship reflects equal and opposite responses in the opponent mechanisms and thus is information that could in principle be available to an observer.

Here we have similarly shown that while a “space” may provide a useful framework for understanding the neural representation of facial expressions, observers do not in fact experience expressions in way that allows them to readily perceive the “spatial” relationships between different expressions. In particular, we found that observers are very poor at judging which pairs of faces are complementary to each other, even though the special nature of these anti-faces has been found to underlie a number of phenomena in face perception. They are inconsistent in their own judgments, and also with each other.

Among specific expressions, it appears that observers were most successful in creating a 180° opposite when the reference was happy (90° with a ~270° opposite), with less accuracy but trending when referencing to anti-happy (270° reference). For some subjects, this had less selectivity; a 45° or 135° reference may have elicited settings clustered somewhere in the anti-happy range. Other faces or anti-faces did not have as much consistency. This may partially be due to the fact that conceptual opposites such as happy vs. sad or even anger are more established than others (what is the opposite of disgust?). This difference may also be due (or in addition) to the simplicity of the concept of reversing a particular expression. The happy manipulation is rather easy; an upside-down smile is similar to a frown, only differing in minute details like the raising cheeks or squinting eyes involved in a genuine smile. And although observers could readily tell a sad face and an upside-down smile apart, when they were shown in isolation the *conceptual* differences did not seem apparent.

One potential limitation that should be addressed is that we do not know the actual dimensions of expression space. It is quite possible that the actual shape of the face space involving two pairs of expressions does not have axes that are orthogonal with respect to each other; instead two expressions may be represented at oblique angles (anti-angry may be more like happy than it is anti-happy. This could affect the presumed angle of intermediate directions in the space (i.e. angles between the two principal axes). But is not in fact a problem for our specific task, since for these oblique axes the faces that are opposite should still lie on opposite sides of the mean.

One account of these results is that even if the stimuli are encoded in a spatial norm-based framework where different expressions are represented by different directions, the expressions themselves are perceived as different qualitative categories rather than different coordinates in the space. That is, anger and fear might be encoded as “apples and oranges” and the visual system may therefore not be designed to explicitly represent how an apple and orange differ, or whether they are more similar to each other than a pear. In that case observers might still be good at judging the “quantities” of a category or their relative magnitude. The second set of experiments addressed this by comparing intensity matching across different expressions. This work was a direct extension of the cross-color contrast matching experiments of Switkes and Crognale (1999) to face perception. We found that observers could readily abstract the intensity independent of the direction for real expressions. However, for anti-expressions this task became much more difficult. Notably these facial configurations don't have a clear perceived expression at all, and it is surprisingly difficult to estimate how the strength of the configuration varies. This could be a flaw of the experiment. Real expressions distort the face with specific muscle patterns that may not have a physically possible opposite (e.g. you can open a mouth to smile but cannot over-close it to produce an anti-smile). However, whether plausible physically, these configurations must at some level play a role in expression coding – at least according to the dominant norm-based accounts. If they did not, then there would be no anti-face to anchor the norm (Webster and Juricevic, 2012).

Suppose each expression were coded as a different object and observers could sense the magnitude of that object. In that case if they had to pick an average of two expressions they might independently set the average intensity and average angle. Our last experiment tested this hypothesis, and surprisingly, it is not what we found. The average of two expressions that were 90° apart in the space was set on a line running directly between the components. This simple linear average must plot closer to the origin and thus have a lower expression intensity than the reference faces. This stands in contrast to the results of experiment 1, which effectively involved the same task but a wider angular separation. Thus one possibility is that observers can perceive spatial variations as long as the stimuli are not too far apart within the space. However, another possibility is that the observer's intuitions in the two cases were different. For example, it may be that people do not even have an intuitive understanding of opposites (in the sense that they can readily visualize them) but can nevertheless compute something like an average even for very large stimulus differences. Future work could address this by varying the actual instructions given to observers or by using the types of limited attention tasks which have been more widely used to explore the perception of summary statistics (Haberman et al., 2009; Whitney et al., 2013).

# IV. General Discussion

These experiments explored the organization of face representations in a conceptual neural face space. Similar to color space models, the appearance of the face can potentially be predicted by its specific position in the space and its relation to the other faces. The two main models describing many details of the space can be contrasted by whether they characterize faces by their absolute position in the space (exemplar-based) or their position relative to a fixed norm (norm-based) (Valentine, 1991). The current body of research appears to more strongly support the existence of a norm. Characteristic of the norm-based model is that as faces deviate farther in space from the norm, they become less like it and more distinct in appearance, whether that means more happy, more female, or more old. The angle of a given face relative to the norm can also be changed, altering the form of the face, e.g. how similar it is to a given "pure" emotion in face space). Similar to color opponency models, as a face becomes less angry it must pass through the norm before it can begin to increase in the anti-angry direction. More evidence for a norm is shown with expression selectivity (Juricevic & Webster, 2012; Skinner & Benton, 2010) as well as aftereffects that are proportional to the adapt strength (Robbins et al., 2007). Although there is evidence that some face properties rely on exemplar coding or narrowband channels (Calder et al., 2008), the majority of findings support representation relative to an average face. There are still some questions about how much other face categories are supported by this model, or how we average and make predictions across face space. The three main experiments conducted add to our knowledge about how face space is laid out.

Experiment I looked at our perception of age and found evidence of an age norm. Age is a unique measure because you can quantify age judgments and examine how a range of test ages is affected by an adaptor. This allows us to more precisely examine the norm vs. exemplar model, as the former relies on broadband neuronal channels which should cause aftereffect at a wider range of ages and would cause all ages to shift in the same direction. Exemplar models predict more localized results. This study found results that strongly supported age relying on norm-based coding. Another piece of evidence is that middle age, as determined by psychophysical measurements for each observer, does not cause significant aftereffects. These results suggest that age perception relies on similar mechanisms as other face categories and appears to be represented by a similar opponent axis.

Experiment II tested how we might neurally integrate two competing images in our visual field. Viewing a neutral and a distorted face in each eye caused rivalry between the two. While initial results suggested that the distorted face was dominant, this effect disappeared or was reduced when more response choices were allowed or when fusion was limited. One possible explanation is that we encounter more faces similar to the norm than highly distinct faces, thus the density of faces is much higher in the center of face space. Because there are many more prototypical representations here, a face must be *very* similar to a known prototype to be put in the same group, while distorted faces are more readily classified as distorted even if physically quite different (Tanaka et al., 2004). Therefore, we have an inherent bias outside of rivalry to classify new faces in intermediate face space as distorted rather than non-distorted. This model also does not seem to be face specific, strengthening its possible links to other modalities. Evidence of an adaptation-based underpinnings is seen both with naïve subjects viewing common objects (Tanaka & Corneille, 2007; Tanaka, Kantner, & Bartlett, 2012) and with novel objects that the observers were trained to discriminate (Kantner & Tanaka, 2012).

These results suggest similarities to results in adaptation and other areas. Models that support a norm consider it to be "unique" rather than just another face, and thus other faces should interact differently depending on their relative position. A distinct face is far from the norm and should more strongly activate the visual system, similar to results seen in adaptation. In addition, rivalry seems to average or integrate the images, causing new distinct percepts. Observers in the third experiment reported seeing the 50% distorted face for a statistically equal length and frequency as the physical stimuli were reported, despite it not actually being presented. The two images were not merely averaged in color and luminance, although that happened at a much lower rate, but observers perceived a high-level integration of the two extremes. The new percept might have not been exactly halfway along the space vector between the extremes, but when given additional response choices, the percept was judged as most similar to the new prototype. This provides evidence that a model that relies on the norm is a better explanation than an exemplar model.

Experiment III looked at our ability to consciously access our own face space and recognize how expressions and expression intensity interrelate. In most cases, observers were unable to use face space calculations to any consistent effect. In the first experiment they simply had to indicate the opposite of a given face, and they made settings that varied greatly between repetitions of the same trial (high variance). Also, different observers had little agreement with each other except in specific areas. While our face processing relies on a norm-based space, we don't seem to be aware of how two faces may relate to each other, especially when they are far apart in face space (180°). This may be similar to color spaces where we seem to have difficulty making opposite hue settings (Webster et al., 2013). One thing to note is that it is difficult, especially with observers who are naïve to color science, to imagine an average of red and green, or blue and yellow. It is not intuitive that the average is gray or that are opponent colors. With the comparison between two opposed faces is also difficult to see that they average to a neutral, expressionless face. Predicting the anti-face requires knowing the face structure changes needed to go to neutral, and then extrapolating and continuing those changes an equal distance. With the happy expression, it was perhaps easier to conceptualize, but it is more difficult to imagine the shape opposite of other expressions, and no meaningful settings could be made.

A related question is to look at our ability to recognize changes in intensity that occur in the face space radial direction. We first tested whether observers could recognize the relative intensities of two expressions or two anti-expressions. They made fairly accurate settings, matching high (low) intensity reference and test faces between paired expressions. The observers did not always give the expressions equal strength (as based on FaceGen's scale), for example happy faces needed to be stronger than angry or fearful faces. But what is more important is that the magnitudes were consistent: as fear intensity increased, happy settings increased by a proportionate magnitude and vice versa. More variable results were found when matching anti-expressions with another anti-expression, but observers could still recognize relative intensities. It is important to note that these comparisons are essentially 90° in face space. If we are unable to match expressions across 180°, are we also unable to match intensities of faces at that distance? When the same task was completed with face/anti-face pairs, the observers again made inconsistent judgments with no monotonic change in the test. One thing to note is that the opposite color requires a transformation passing through the norm. It may be that this "barrier" prevents direct comparisons to be made. This might be analogous to color models: again, gray is the average of colors on opposite sides of a color axis, but this is not conceptually apparent. Similarly, finding the opposite expression or its intensity requires a computation that we are perhaps not primed to make. Our ability to estimate or anticipate changes in face space appears to be strong but not perfect, especially across far distances.

These experiments collectively explore our ability to navigate our own representation of face space. The relationships within this space are created through constant adaptation to faces we encounter, thus the layout and norms differ between individuals. Nevertheless, face spaces in general rely on similar rules and constraints. Categories such as expression, distinctiveness, and age rely on a representation of a norm. This effect parallels many parts of our color vision, but also seems to be not entirely face-specific (Rhodes, Jeffery, Boeing, & Calder, 2013). But while we access face space while interpreting faces, we don't have full access to this information to make conscious examinations of the layout of face space.

Further information about the relationships in face space can help answer many questions. First, the structure of face space is being shown incrementally by many studies. Much information can be provided by determining the axes used in face space and how they interact, as well as information about the uniformity of the space and how norms are created. As we are constantly adapting, our norm is constantly shifting subtly, resulting in faces appearing more or less similar to it. This can be implicated in other-race, other-gender, or other-age effects resulting from our experience. But experience with a new pool of faces may gradually shift these biases, increasing the ability to correctly classify members of the other group (Tanaka et al., 2004). If we understand more about this process, we may be able to predict or train for greater discrimination of faces. This may also have use in areas like facial recognition. If a system is designed to analyze faces as quickly as possible, it needs to efficiently determine what makes this particular face distinct by its vector in face space. This process may be facilitated by having access to a comparison face, namely a neutral face that is the norm of all the faces that the system can reasonably expect to encounter. It might also be possible to create different norm definitions based upon the dominant ethnicity, age or gender in an area, increasing sensitivity for these faces while still recognizing faces that are very different. A facial recognition system could also potentially be adaptive and could recalibrate the norm constantly, or account for the common problem of recognizing the same identity even if the appearance has slightly been changed. Finally, by testing if more features besides expression, identity, age, etc. fall in a face space with positions relative to a norm, we may determine if any similarities can be predicted compared to other neural processes like color. Of interest in the future is to learning more about the extent that the identity of a face is similar to other attributes, and in particular to compare to the current findings with expression. Identity has been shown to use a face space (Leopold et al., 2001; Yankouskaya, Booth, & Humphreys, 2012), and has many similarities known, but its appearance when the angular direction changes is not apparent. Changing identities also may not be completely perceptually independent; age, absolute distinctiveness, or luminance may not remain constant in an angular direction, particularly if axes are oblique as is often assumed. Nevertheless, identity appears to be a "real" category and not just the sum of its parts, and can be adapted independent of expression (Skinner & Benton, 2012) and potentially age, thus allowing us to do things like recognize a person that we have not seen for some time. The current findings support the likelihood that new findings will show independent axes in face space, but all of which are refernced to a constantly calibrated norm.

## References

Bannerman, R. L., Milders, M., De Gelder, B., & Sahraie, A. (2008). Influence of emotional facial expressions on binocular rivalry. *Ophthalmic and Physiological Optics*, *28*(4), 317–326.

Blake, R., & Wilson, H. (2011). Binocular vision. *Vision Research 50th Anniversary Issue: Part 1*, *51*(7), 754–770. doi:10.1016/j.visres.2010.10.009

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

Breese, B. B. (1899). On inhibition. *The Psychological Review: Monograph Supplements*, *3*(1), i–65. doi:10.1037/h0092990

Breese, B. B. (1909). Binocular rivalry. *Psychological Review*, *16*(6), 410–415. doi:10.1037/h0075805

Burt, D. M., & Perrett, D. I. (1995). Perception of Age in Adult Caucasian Male Faces: Computer Graphic Manipulation of Shape and Colour Information. *Proceedings: Biological Sciences*, *259*(1355), 137–143. doi:10.2307/50147

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

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

Chong, S. C., & Blake, R. (2006). Exogenous attention and endogenous attention influence initial dominance in binocular rivalry. *Vision Research*, *46*(11), 1794–1803.

Darwin, C. R. (1872). *The expression of the emotions in man and animals*. London: John Murray.

De Fockert, J. W., & Gautrey, B. (2013). Greater visual averaging of face identity for own-gender faces. *Psychonomic Bulletin & Review*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-02839-001&site=ehost-live

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

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., … Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712–717.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88.

Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception*, *39*(7), 931–941.

Enlow, D. H. (1982). *Handbook of facial growth* (2nd ed.). Philadelphia: Saunders.

Fang, F., & He, S. (2005). Viewer-Centered Object Representation in the Human Visual System Revealed by Viewpoint Aftereffects. *Neuron*, *45*(5), 793–800. doi:10.1016/j.neuron.2005.01.037

Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research*, *1127*(1), 80–89.

Gao, X., Maurer, D., & Nishimura, M. (2010). Similarities and differences in the perceptual structure of facial expressions of children and adults. *Journal of Experimental Child Psychology*, *105*(1-2), 98–115.

George, P. A., & Hole, G. J. (2000). The role of spatial and surface cues in the age-processing of unfamiliar faces. *Visual Cognition*, *7*(4), 485–509.

Gray, K. L. H., Adams, W. J., Hedger, N., Newton, K. E., & Garner, M. (2013). Faces and awareness: Low-level, not emotional factors determine perceptual dominance. *Emotion*, *13*(3), 537–544.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2009-21237-001&site=ehost-live

Hancock, S., Whitney, D., & Andrews, T. J. (2008). The initial interactions underlying binocular rivalry require visual awareness. *Journal of Vision*, *8*(1), 3–3. doi:10.1167/8.1.3

Jeffery, L., McKone, E., Haynes, R., Firth, E., Pellicano, E., & Rhodes, G. (2010). Four-to-six-year-old children use norm-based coding in face-space. *Journal of Vision*, *10*(5). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-09797-001&site=ehost-live

Jeffery, L., & Rhodes, G. (2011). Insights into the development of face recognition mechanisms revealed by face aftereffects. *British Journal of Psychology*, *102*(4), 799–815.

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

Juricevic, I., & Webster, M. A. (2012). Selectivity of face aftereffects for expressions and anti-expressions. *Frontiers in Psychology*, *3*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-27673-001&site=ehost-live

Kang, M.-S., & Blake, R. (2010). What causes alternations in dominance during binocular rivalry? *Attention, Perception, & Psychophysics*, *72*(1), 179–186.

Kantner, J., & Tanaka, J. W. (2012). Experience produces the atypicality bias in object perception. *Perception*, *41*(5), 556–568.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(ECVP Abstract Supplement).

Knapen, T., Kanai, R., Brascamp, J., van Boxtel, J., & van Ee, R. (2007). Distance in feature space determines exclusivity in visual rivalry. *Vision research*, *47*(26), 3269–3275. doi:10.1016/j.visres.2007.09.005

Lai, M., Oruç, I., & Barton, J. J. S. (2012). Facial age after-effects show partial identity invariance and transfer from hands to faces. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *48*(4), 477–486.

Lai, M., Oruç, I., & Barton, J. J. S. (2013). The role of skin texture and facial shape in representations of age and identity. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *49*(1), 252–265.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. doi:10.1080/02699930903485076

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

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

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94.

Leppänen, J. M., Richmond, J., Vogel-Farley, V. K., Moulson, M. C., & Nelson, C. A. (2009). Categorical Representation of Facial Expressions in the Infant Brain. *Infancy*, *14*(3), 346–362. doi:10.1080/15250000902839393

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363–384.

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

McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *17*(2), 211–224.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments & Computers*, *36*(4), 630–633.

Mueller, T. J., & Blake, R. (1989). A fresh look at the temporal dynamics of binocular rivalry. *Biological Cybernetics*, *61*(3), 223–232. doi:10.1007/BF00198769

Munsell, A. H. (1905). *A Color Notation*. G. H. Ellis Company. Retrieved from http://books.google.com/books?id=PgcCAAAAYAAJ

Nishimura, M., Doyle, J., Humphreys, K., & Behrmann, M. (2010). Probing the face-space of individuals with prosopagnosia. *Neuropsychologia*, *48*(6), 1828–1841.

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

O'Toole, A. J., Vetter, T., & Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing. *Vision Research*, *39*(18), 3145–3155.

Oruç, I., & Barton, J. J. S. (2011). Adaptation improves discrimination of face identity. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1718), 2591–2597. doi:10.1098/rspb.2010.2480

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

Peltola, M. J., Leppänen, J. M., Mäki, S., & Hietanen, J. K. (2009). Emergence of enhanced attention to fearful faces between 5 and 7 months of age. *Social Cognitive and Affective Neuroscience*, *4*(2), 134–142.

Principe, C. P., & Langlois, J. H. (2012). Shifting the prototype: Experience with faces influences affective and attractiveness preferences. *Social Cognition*, *30*(1), 109–120.

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

Rhodes, G., Jeffery, L., Boeing, A., & Calder, A. J. (2013). Visual coding of human bodies: Perceptual aftereffects reveal norm-based, opponent coding of body identity. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 313–317. doi:10.1037/a0031568

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

Rhodes, G., & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In *The Oxford handbook of face perception* (pp. 263–286).

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

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

Rodríguez, J., Bortfeld, H., Rudomín, I., Hernández, B., & Gutiérrez-Osuna, R. (2009). The reverse-caricature effect revisited: Familiarization with frontal facial caricatures improves veridical face recognition. *Applied Cognitive Psychology*, *23*(5), 733–742.

Ross, D. A., Deroche, M., & Palmeri, T. J. (2013). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin & Review*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-18103-001&site=ehost-live

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, *35*(6), 749–759.

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

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

Skinner, A. L., & Benton, C. P. (2012). The expressions of strangers: Our identity-independent representation of facial expression. *Journal of Vision*, *12*(2). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-05245-001&site=ehost-live

Stanley, J., Carter, O., Forte, J., & Martinez-Conde, S. (2011). Color and luminance influence, but can not explain, binocular rivalry onset bias. *PLoS ONE*, *6*(5). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2011-09767-001&site=ehost-live

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

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

Switkes, E., & Crognale, M. A. (1999). Comparison of color and luminance contrast: Apples versus oranges? *Vision Research*, *39*(10), 1823–1831.

Tanaka, J., & Corneille, O. (2007). Typicality effects in face and object perception: Further evidence for the attractor field model. *Perception & Psychophysics*, *69*(4), 619–627. doi:10.3758/BF03193919

Tanaka, J., Giles, M., Kremen, S., & Simon, V. (1998). Mapping attractor fields in face space: the atypicality bias in face recognition. *Cognition*, *68*(3), 199–220. doi:10.1016/S0010-0277(98)00048-1

Tanaka, J. W., Kantner, J., & Bartlett, M. (2012). How category structure influences the perception of object similarity: The atypicality bias. *Frontiers in Psychology*, *3*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-26226-001&site=ehost-live

Tanaka, J. W., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition*, *93*(1), B1–B9.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *43A*(2), 161–204. doi:10.1080/14640749108400966

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *44A*(4), 671–703.

Van Ee, R. (2005). Dynamics of perceptual bi-stability for stereoscopic slant rivalry and a comparison with grating, house-face, and Necker cube rivalry. *Vision Research*, *45*(1), 29–40.

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

Webster, J., Nolan, C., Sternberg, H., Kay, P., & Webster, M. A. (2013). Numbering by color: the perception of summary statistics in color distributions. *Journal of Vision*, *13*(9), 1165–1165. doi:10.1167/13.9.1165

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

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

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

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

Whitney, D., Haberman, J., & Sweeny, T. D. (2013). From textures to crowds: multiple levels of summary statistical perception. In J. S. Werner & L. M. Chalupa (Eds.), *The New Visual Neurosciences*. The MIT Press.

Wolfe, J. M. (1984). Reversing ocular dominance and suppression in a single flash. *Vision Research*, *24*(5), 471–478.

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, *114*(1), 101–114.

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

Yankouskaya, A., Booth, D. A., & Humphreys, G. (2012). Interactions between facial emotion and identity in face processing: Evidence based on redundancy gains. *Attention, Perception, & Psychophysics*, *74*(8), 1692–1711.

Yoon, K. L., Hong, S. W., Joormann, J., & Kang, P. (2009). Perception of facial expressions of emotion during binocular rivalry. *Emotion*, *9*(2), 172–182.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*, *63*(3), 271–313. doi:10.1016/S0010-0277(97)00003-6

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, *17*(6), 263–271. doi:10.1016/j.tics.2013.04.004

Yu, K., & Blake, R. (1992). Do recognizable figures enjoy an advantage in binocular rivalry? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1158–1173.

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

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