Plasticity of face perception in the human visual system

A DISSERTATION SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Xinyu Liu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisor: Dr. Stephen A. Engel

August 2022

Abstract

When our surrounding perceptual environment changes, our sensory systems change in tandem to help us process information efficiently and improve perception. In the visual system, adaptation and perceptual learning are examples of this plasticity. This dissertation presents three studies investigating how adaptation and perceptual learning improve our ability to see faces. Study 1 explored whether higher-level adaptation to faces can help overcome illusory percepts caused by adaptation at lower-level visual stages. We found that the visual system can adapt to distorted faces caused by adaptation at lower-level visual areas through repeated viewing; we termed this phenomenon "meta-adaptation". Meta-adaptation may be a general strategy to correct negative consequences, e.g. illusory percepts, of lower-level adaptation. Study 2 investigated whether perceptual learning could help observers form mental representations of groups of faces. We showed that 3 days of perceptual learning made faces that initially appeared similar more dissimilar; clustering analysis further suggested that training strengthened group-level representations for the trained faces. In study 3, we more directly tested whether perceptual learning could form group-level representations, by measuring transfer of face adaptation effects between two trained groups of faces. Results showed learning caused a decrease in transfer of adaptation between the trained groups, suggesting that after perceptual learning observers were able to develop more distinct representations for each group. Together, these findings help us better understand how face

perception can be influenced by past viewing history and training; the concepts of meta-adaptation and group-level representations could generalize to other domains of visual perception, and other sensory modalities

Table of Contents

List of Figures

Chapter 1

A brief review on visual adaptation and face perception

1.1 Introduction and overview of this dissertation

Our perception of the world is ever-changing, and depends highly on what we have seen previously, and the surrounding perceptual environment. Visual adaptation occurs as neurons adjust their sensitivity and responsiveness in response to the current environment, and often results in changes of perception (Webster, 2011). Our visual system can adapt to many things; for example our perception of brightness (e.g. Whittle, 1992), contrast level (e.g. Blakemore & Campbell, 1969; Ohzawa, Sclar, & Freeman, 1985), orientation (Gibson & Radner, 1937), viewpoint of an object in a three-dimensional space (Fang & He, 2005), shape of a face (see review Webster & Macleod, 2011) and shape of a body part (Kovács et al., 2006), all depend on what we have seen previously.

This dissertation includes three studies that investigated the plasticity of human face perception. In these studies, we used behavioral methods, including visual adaptation, visual search and perceptual dissimilarity to show how viewing history and training affect face perception, and to reveal functional benefits of such plasticity. We developed a behavioral paradigm to show that higher-level face adaptation can correct perceptual errors arising from lower-level adaptation (Chapter 2). We also showed that multi-day perceptual training on groups of faces that initially appeared similar could strengthen the group-level representations of trained faces (Chapter 3); and we presented more direct behavioral evidence for trained group-level representations by measuring transfer of face adaptation effects across groups of faces (Chapter 4).

Visual adaptation played an important role in all 3 studies in this dissertation, therefore in this chapter, I will first provide a general review on visual adaptation (and face adaptation), including its behavioral effects and physiological changes following adaptation, and a general overview of face encoding and representation in the human visual system.

1.2 Visual adaptation: behavioral effects and physiological changes

Perceptual change following adaptation is classically measured as an increase in detection thresholds: for example, after adaptation to a high contrast grating of a particular spatial frequency, it is harder to detect gratings that are similar to the adaptor grating, whereas the detectability of a quite different spatial frequency grating is unaffected (Blakemore & Campbell, 1969).

Many visual adaptation processes can be modeled by gain control mechanisms reducing the responsiveness of adapted neurons. Such gain control can be linear or non-linear (i.e. subtractive or divisive), but essentially this means that cortical neurons that adapt lose their sensitivity across all stimulus intensity levels, which results in a rightward shift in those neurons' stimulus-response curves (Movshon & Lennie, 1979). This change in sensitivity is usually limited to test stimuli similar the adaptor, e.g. adaptation to a grating is orientation and spatial-frequency, or 'pattern', selective.

A loss in sensitivity of adapted neurons can cause various neural and behavioral changes: First, because these neurons are less sensitive (less responsive) to the adapting stimulus, detection thresholds will rise. Second, because the adapting neurons' responsivity curve shifts horizontally, this effectively re-centers neurons' limited dynamic range of output to around the intensity level of the pervasive stimulus (see review by Kohn, 2007). This means adapting neurons now have more dynamic range to encode small differences in intensity levels that are close to that of the adapting stimulus.

In addition to altering sensitivity and tuning profiles, adaptation can potentially modulate suppressive normalization signals from surrounding neurons (see review by Solomon & Kohn, 2014). This points to a complex set of mechanisms that govern visual adaptation. Interestingly, one particular model (see review by Heeger & Carandini, 2012; also see Schwartz & Simoncelli, 2001) of sensory gain control proposes that neural adaptation can itself be modeled as a divisive normalization process. This model is useful in further explaining the functional roles of neural adaptation. More specifically, adaptation can eliminate statistical dependencies that are prevalent in natural stimuli (e.g. shared features, see Turk & Pentland, 1991), and to increase efficiency in representing stimuli (Barlow, 1961; Laughlin, 1981).

Adaptation has been used by psychophysicists as a tool for 'dissecting' the neural mechanisms that our visual system uses for encoding stimuli. This is because, as mentioned above, adaptation effects, both perceptual and physiological, are often specific. That is, adaptation often alters the responsiveness and/or tuning profiles of a specific group of neurons. As a result of this selectivity, psychophysicists can effectively target certain population of neurons (sometimes termed "channels") and identify properties of them by observing perceptual changes following adaptation (see review Kohn, 2007).

1.3 Face adaptation

Webster & Maclin (1999) first found that one's perception of the shape of a face depends on the shape of faces one has been exposed to previously. Vision adapts to different facial features in addition to shape, ranging from expression (e.g. Webster, Kaping, Mizokami, & Duhamel, 2004), to gender (Webster et al., 2004) , race (Rhodes, Watson, Jeffery, & Clifford, 2010) and other facial attributes.

Norm-based encoding models (e.g. Leopold, Bondar, & Giese, 2006; Leopold, O'Toole, Vetter, & Blanz, 2001) explain perceptual changes due to adaptation by proposing that it shifts the 'norm' for faces towards the adapting face. In norm-based face models, all faces are represented as differences/deviations from a special face called the "norm face". The particular norm used in most models is an average of all exemplars seen (e.g. Rhodes & Jeffery, 2006), and is usually weighted over time so that recently seen faces have greater influence. Accordingly, exposure to a particular face shifts the norm towards that face. This shift in turn causes a previously normal-appearing face, which used to have only a very small difference from the norm, to appear 'opposite' of the adapting face, since it now has a difference from the norm in the opposite direction from the adapter-norm difference. For example, prolonged exposure to a masculine face causes a physically gender-neutral face to appear more feminine (Webster et al., 2004).

Face adaptation has been hypothesized serve important functional roles. For instance, it may remove 'irrelevant information' for face identification, similar to discounting the ambient spectrum when seeing color or discounting viewpoint when recognizing objects. Specifically, when viewing faces, our main perceptual task might be to estimate the identity given the image of a face. Although many variables could influence the perception of a face (e.g. the overall shape of a face, facial expressions, eye position, even skin texture and skin tone), it would be helpful to discount any variable that does not help identify facial identity (Kersten & Yuille, 2003). Chapter 2 will examine whether adaptive norm shifts can help discount distortions in faces produced by lower-level adaptation.

As mentioned earlier, adaptation is a useful behavioral method to infer neural coding, given that effects of adaptation are often specific to the group of neurons that encode the adapting stimulus. This logic applies to face adaptation: When face adaptation aftereffects transfer between faces, it implies shared encoding/representation between these faces. For example, face identity aftereffects was found to transfer across facial expressions (Fox, Oruç, & Barton, 2008). This result implies that faces of different expressions but belonging to the same identity are represented similarly by our visual system. Chapter 4 will take advantage of this logic to test for changes in the representations of groups of faces.

Chapter 2

Higher-level ‘Meta’-Adaptation Mitigates Visual Distortions Produced by Lower-level Adaptation

When our environment changes, our sensory systems change in tandem, in order to keep us perceiving well. In the visual system, this adaptation produces changes in neural response at early stages of processing that are large, common, and likely improve the efficiency of encoding (for reviews see Clifford et al., 2007; Kohn, 2007; Webster, 2015). Nevertheless, adaptation can also lead to misperceptions of the world, due to what is known as the “coding catastrophe” (Fairhall, Lewen, Bialek, & de Ruyter van Steveninck, 2001; Schwartz, Hsu, & Dayan, 2007). Both perceptual and neurophysiological work have shown that adaptation arising in early visual areas propagates along the visual stream, changing responses in later stages to a fixed stimulus (for examples see Dhruv & Carandini, 2014; Patterson, Wissig, & Kohn, 2014; Xu, Dayan, Lipkin, & Qian, 2008). These altered responses are sometimes interpreted as arising from a change in the stimulus, rather than a change in the neural response pattern (or code), which produces incorrect, “illusory” percepts (Schwartz et al., 2007).

For example, a standard illusory misperception— the tilt aftereffect—is produced when adaptation to a high-contrast pattern of tilted stripes, for example oriented at 20 degrees, is followed by presentation of a physically vertical pattern, which then appears tilted in the opposite orientation of the adaptor pattern (Gibson & Radner, 1937). Adaptation to high contrast patterns is known to cause a reduction in neural responsiveness in the primary visual cortex (V1) (e.g. Fang, Murray, Kersten, & He, 2005; Movshon & Lennie, 1979), particularly for neurons that prefer orientations similar to the adaptor. This adaptation causes presentation of the vertical pattern to produce a distribution of responses, across neurons in early visual cortex, that is changed from its un-adapted state; in this example its center of mass would be shifted away from 20 degrees. The accepted interpretation of the tilt aftereffect is that neurons in later visual areas respond to this altered distribution of input from V1 as if it were generated by a physical stimulus that was shifted away from 20 deg, producing the illusory tilt (Fairhall et al., 2001; Schwartz et al., 2007).

Adaptation can also occur at later processing stages, however. Many high-level percepts encoded later in processing change following adaptation, including the shape of a face (Webster & Maclin, 1999), the specific viewpoint for observing an object (Fang & He, 2005) and even the gender of a hand (Kovács et al., 2006). Given feedback connections in cortex and uncertainty regarding the neural bases of perceptual measures, strictly defining earlier and later stages of processing can be challenging. Here, we refer to later stages of processing as those that receive feed-forward input from earlier stages, and lower-level and higher-level adaptation as changes in response properties of neurons at earlier and later stages of processing, respectively. When direct neural evidence is lacking, investigators take advantage of the hierarchical nature of the visual pathways, and define lower-level adaptation as that affecting representations of features, and higher-level adaptation as that affecting percepts based upon integrating those features (e.g. see Xu et al., 2008).

Some higher-level adaptation can be explained as shifts in perceptual norms: For example, prolonged exposure to a masculine face causes a physically gender-neutral face to appear more feminine (Webster, Kaping, Mizokami, & Duhamel, 2004). Norm-based encoding accounts for this effect by proposing that adapting to a masculine face shifts a norm for faces towards male faces. This shift in turn causes the previously neutral face to appear more feminine, since it is encoded relative to the norm. Norm-based codes are also believed to play an important role in color perception (Webster & Leonard, 2008), perception of blur (Elliott, Georgeson, & Webster, 2011), body identity perception (Rhodes, Jeffery, Boeing, & Calder, 2013), and other perceptual qualities.

Can this higher-level adaptation help overcome the negative consequences of lower-level adaptation? We hypothesized that norm-based encoding may help resolve the coding catastrophe if adaptation propagates along the visual stream, and later-stage visual areas adapt to the altered input received from earlier stages. While past work has made clear that lower-level adaptation can influence processing at later stages (e.g. Dhruv & Carandini, 2014; Patterson et al., 2014), the extent to which later areas can adapt to the altered input they receive remains widely unexplored (though for one example, see Webster & Mollon,1995).

We tested our hypothesis with lower-level adaptation that made a physically normal face appear distorted. We expected repeated viewing of this distorted face to produce adaptation at later stages of face processing, an example of what we term “meta-adaptation”. The meta-adaptation should shift perceptual norms towards the adapter- that is towards the distorted face- and so cause it to appear more normal (Figure 1). Our results supported our hypothesis, documenting meta-adaptation, showing its build-up over time, and advancing understanding of how adaptation at different levels can work together to optimize behavior (Webster & Mollon, 1995, 1997).

Experiment 1

Participants

15 observers participated in Experiment 1. Sample size was determined from the prior literature on face adaptation (e.g. Leopold, O'Toole, Vetter, & Blanz, 2001; Webster, Kaping, Mizukami, & Duhamel, 2004), as well as an unpublished pilot version of the experiment. Data collection stopped when sample size reached 15. Experimental procedures were approved by the University of Minnesota Institutional Review Board (IRB 1006M84457), and conformed to the Helsinki Declarations.

Stimuli

To produce lower-level adaptation that distorted a face, observers viewed a contrast-reversing checkerboard pattern and a familiar face, in interleaved fashion (Figure 2.1a). The checkerboard patterns were placed near the eye position of the familiar face, to produce shifts in the perceived location of the eyes, making them appear further away from each other. The patches of checkerboard pattern were 0.5 deg tall and wide, and were presented at 95% Michelson contrast, with checks alternating between positive and negative contrast at 8 Hz. Observers' viewing distance was maintained at 52 cm.

The familiar face used in this experiment was of former U.S president Barack Obama,15 deg by 9.6 deg, displayed at the center of a video screen. Test images with different inter-eye distances were created that differed by steps of 0.23 deg visual angle. Mean luminance was 42 candela/m^2, and stimuli were delivered in Matlab using the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

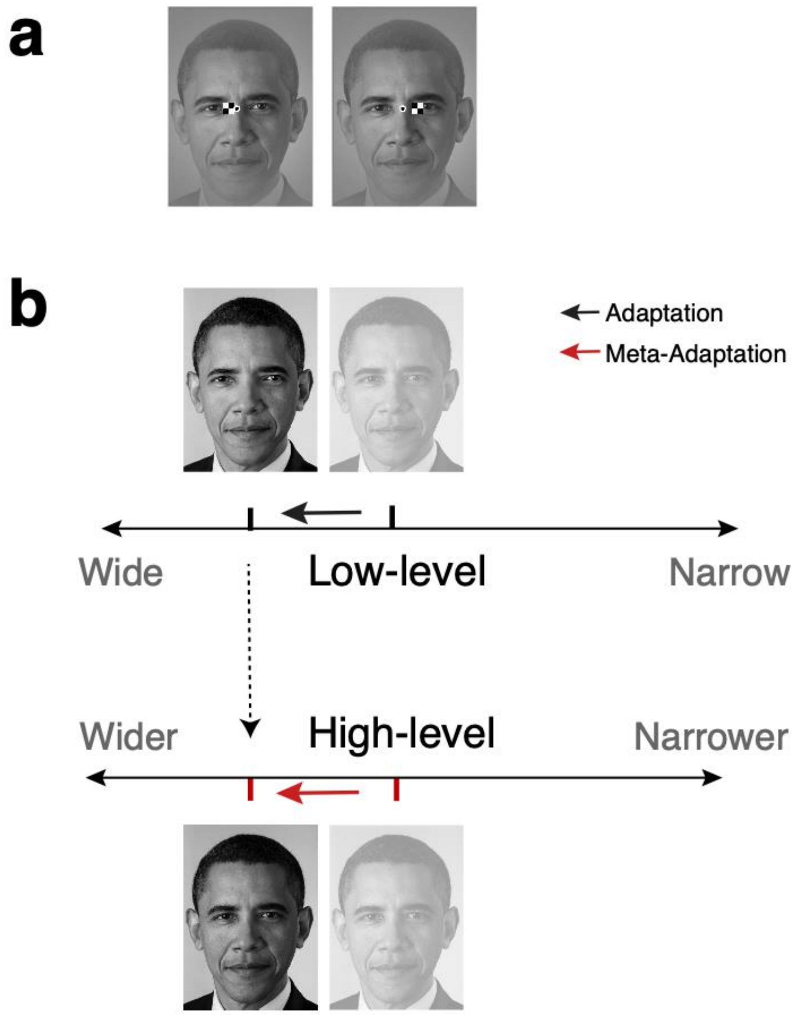

Figure 2.1 Meta-adaptation
a) To produce lower-level adaptation, high contrast checkerboard patterns were placed at locations near the eyes of a subsequently presented undistorted face. Note that the face beneath the checkerboard was not presented during adaptation; it is here to illustrate the relative retinal locations of the checkerboard and the later presented face. **b)** This lower-level adaptation (black arrow) caused the eyes of the subsequently presented face to be represented as shifted outward, wider than normal, due to changes in response of neurons at early stages of the visual system (black tick marks). The low-level representation, altered by adaptation, is inherited by later stages of processing (vertical, dotted black arrow). Meta-adaptation is higher level adaptation driven by this altered input (red arrow). Because the higher-level adaptation shifts the perceptual norm (red tick marks), the face should appear more normal as it is viewed for a prolonged duration, correcting the effects of the coding catastrophe. The face images in b) represent the content of lower-level and higher-level visual processing stages.

Procedure

The observers' task consisted of judging a test face image centered 1.87 deg above fixation. Observers judged whether this face had narrower or wider eyes than an undistorted face image. Test faces were presented for 0.3 sec. The position of the eyes of the test face was controlled by a one-up-one-down staircase procedure, based on the observers' response, which converged to an inter-eye distance that was equally likely to be judged as too narrow and too wide. We estimated this point of subjective equality (PSE) by fitting psychometric functions to the responses, as described below.

Observers participated in up to two practice sessions before the main experiment. The practice sessions began with familiarization to the undistorted face; it was shown for 0.5 sec, followed by a 1 second presentation of a uniform mean gray screen, for a total of 60 seconds. Observers were instructed to closely examine the face's features.

Next, observers practiced the behavioral task in three training blocks, each consisting of 30 trials. Practice session results were used to determine whether the observer had internalized an unbiased and consistent estimate of the inter-eye distance of the undistorted face. Observers were invited to participate in the main experiment if their PSE across blocks varied on average by one level of the inter-eye distance shift used to create the stimuli, i.e. the variation was smaller than 0.23 degrees visual angle. Otherwise they participated in an additional practice session.

In the main session, we measured observers' perceptual norm of a test face, prior to and following meta-adaptation. In pre-adapt trials, which were identical to the practice sessions, observers saw a test face presented above the center of the screen for 0.3 seconds and had 2 seconds to determine whether the face presented had narrower or wider inter-eye distance than a normal face. Each observer completed 30 trials.

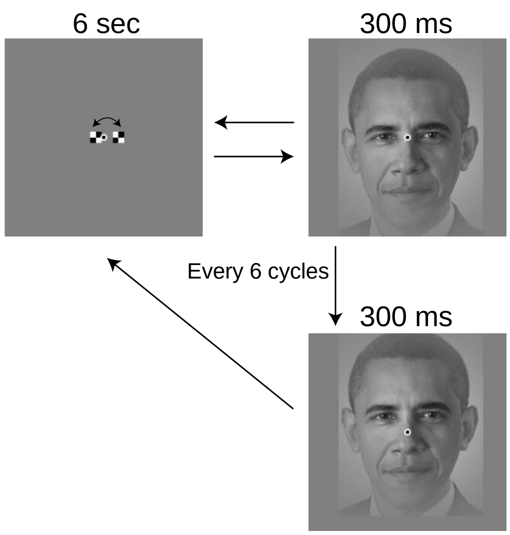

Figure 2.2 Experimental procedures.
Observers initially viewed 2 min of contrast-reversing checkerboard. Patches to the left and right of fixation alternated in time, with each shown for 1 sec. After 2 min, a normal face was shown at the center of the screen briefly for 300 ms, followed by 6 sec of "top-up" adaptation to the checkerboards. This top-up checkerboard-face cycle was repeated 6 times (over 38 sec), and then a test face was shown in the upper visual field, where its eye positions were relatively

The main session began with 2 minutes of contrast adaptation, in which observers viewed the contrast-reversing checkerboard patterns that were placed near fixation. This contrast adaptation caused the eyes of the undistorted face to appear shifted outwards, away from the contrast-reversing pattern. The likely mechanism for this is that adaptation caused neurons that represent the adapted location to decrease their responsiveness, which resulted in an overall shift in the neural population response distribution to the eyes (i.e. amount of neural response as a function of receptive field location) away from the adaptor (Dhruv & Carandini, 2014). This in turn caused the perceived location of the eyes to shift away from the adapted location.

After the initial 2-minute contrast adaptation, observers viewed 0.3 sec of the physically undistorted face (which appeared distorted from the contrast adaptation aftereffect), followed by 6 seconds "top-up" contrast adaptation (Figure 2). This face checkerboard-top-up cycle repeated 15 times before the first test face presentation, in order to attempt to produce strong initial meta-adaptation to the distorted face. To ensure observers were attending to the face presented, one eye of the face occasionally would "wink" (movement in the eye lid), and observers were instructed to press a button upon detecting this movement.

After the 15th checkerboard-face cycle, observers again judged a test face. The test face was presented above fixation to ensure that it was not affected by the low-level contrast adaptation, which should be tied to its retinotopic location. Each test presentation was followed by six additional repetitions of the checkerboard-face top-up cycle. This procedure continued for 30 trials, and we again used a staircase to estimate PSE. A pilot experiment showed that this procedure can measure strong norm shifts when subjects viewed physically distorted faces (i.e. traditional face adaptation). The main session lasted around 45 minutes.

Analysis

To estimate PSEs, we fit a psychometric function to observers' responses, a Weibull function, which has the general form:

$$y = 1 - e^{-(\frac{x}{\alpha})^{\beta}}$$

where *y* is the fraction of "too narrow" responses, *x* is the physical inter-eye distance, and α and β are free parameters estimated by the fitting procedure. We used a maximum likelihood fitting, and the PSE was estimated as the physical inter-eye distance that produced 50% "too narrow" responses in the fit function.

Results

To produce adaptation at early stages of processing, observers viewed contrast-reversing checkerboard patterns, placed at visual field locations just central to where the eyes of our test face would appear (Figures 2.1 & 2.2). Adaptation to contrast patterns such as these is known to produce effects in early visual cortex (e.g. Movshon & Lennie, 1979). Perceptually, adaptation to the checkerboard produced a repulsive spatial aftereffect (Dhruv & Carandini, 2014), causing the eyes of a subsequently presented face to appear shifted outward. A pilot experiment verified that this paradigm produced perceptual distortions that were specific to the adapted retinal locations, a sign of their arising relatively early in visual processing.

Can higher levels of the visual system adapt to this face distorted by lower-level adaptation? We measured higher-level face adaptation using a standard method, asking observers to judge the appearance of a face placed at a different location in the visual field from a repeatedly viewed adapting face. The spatial transfer, as well as other past work, indicates that this paradigm depends upon adaptation at later stages of processing(e.g. Leopold, O'Toole, Vetter, & Blanz, 2001). We measured observers' point of subjective equality (PSE) for eye separation, which is the face that appeared to have eyes neither too wide or too close together. We report PSEs as percentage change scores relative to the true normal inter-eye distance (computed by first subtracting and then dividing the raw scores by 1.5 deg, the distance in the undistorted image).

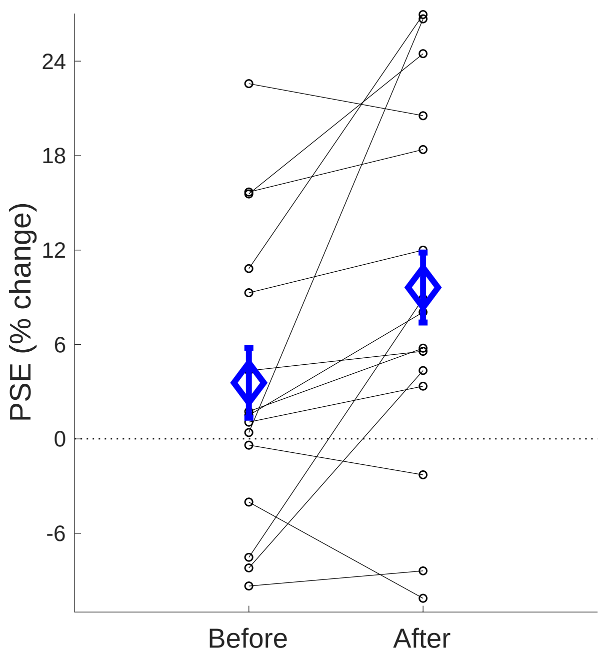

Figure 2.3 Results from Exp. 1.

Individual observers' points of subjective equality (inter-eye distance that appeared neither too close nor too far apart), are plotted with the group mean, measured both before and after adaptation. Error bars represent the standard error of the mean difference score (as appropriate for a within-subjects design, Loftus & Masson, 1994). The dotted line represents 0% change in PSE, i.e. the physical inter-eye distance of the undistorted face. Positive change corresponds to a shift of PSE to a face that has wider eyes.

Face perception showed clear evidence of adaptation following exposure to the face distorted by low-level adaptation. Observers' mean PSE shifted significantly following exposure to the distorted face, compared to a baseline condition measured prior to viewing ($t(14)=2.72$, $p=0.008<0.01$, one-tailed, 95% CI [3.91% $+\infty$%], $d=0.70$). Figure 3 plots individual observer's PSEs before and after exposure to the distorted face, as well as the group mean. The mean PSE increased by about 6%, and 12 out of 15 observers showed trends in the predicted direction, i.e. towards a face with eyes physically farther apart.

The PSE measured the physical face whose eyes appear to be the "correct" distance apart, which is one definition of a perceptual norm. This norm shifted towards faces with eyes farther apart, the same direction as the shift that the lower-level adaptation produced. We conclude that the norm shift made the faces distorted by lower-level adaptation appear more normal; that is, it effectively corrected some of the distortions in appearance produced by the low-level adaptation.

Consistent with this interpretation, many observers verbally reported that the physically undistorted face they saw at fixation, after adapting to the contrast patterns, initially appeared quite wide-eyed, and then later appeared more normal. Some believed that the face physically changed over successive trials.

Experiment 2

The goal of Experiment 2 was to measure this change in appearance over time. Experiment 1 was optimized to measure the basic meta-adaptation effect, and so began with almost 2 minutes of exposure to the distorted face before the norm task was first performed. This was meant to strengthen the high-level adaptation effect, but may have prevented observation of its build-up, since adaptation can sometimes asymptote after a few minutes. Experiment 2 eliminated the initial checkerboard-face cycles that were presented before data collection began. All other experimental procedures were identical to Experiment 1.

Participants

15 observers participated in Experiment 2; two of these 15 observers also participated in experiment 1. Sample size was determined so that it matches the sample size of experiment 1. Data collection stopped when sample size reached 15. Experimental procedures were approved by the University of Minnesota Institutional Review Board (IRB 1006M84457), and conformed to the Helsinki Declarations.

Procedure

Trials were identical to Experiment 1; observers again indicated whether a test face presented for 300 ms centered in the upper visual field (1.87 deg above center fixation) had narrower or wider eye-distance than a normal face. Eye distance of the test face was controlled by the same 1-up 1-down staircase procedure. Observers completed 30 “baseline” trials and then began the adaptation protocol, which began with 2 minutes of initial contrast adaptation. Test trials then immediately began, and alternated with six repetitions of the checkerboard-face cycle (see Figure 4). Observers again completed a total of 30 trials.

Analysis

In an initial analysis, data were pooled over the entire time course, identically to how analysis of Experiment 1 was conducted. To analyze the time course of the meta-adaptation effect, data were divided into 4 non-overlapping time bins. There were not enough trials in each bin to allow for fitting of psychometric functions, and so we simply averaged the presented stimuli (the staircase levels) within the bin as a simple estimate of observers' PSEs. We then tested for a linear trend in the PSE's across bins. To confirm that simple averaging of the staircase levels is a reasonable estimate of PSE, we performed a Monte-Carlo simulation of our experiment with a model observer responding with PSEs that increased linearly over time. The PSEs obtained from the simple estimate fell on a line that lay close to the model observer's PSEs, and did not reliably differ from them.

Results

An initial analysis pooled data across the entire time course as in Experiment 1, and showed a similar meta-adaptation effect. Observers' adapted PSE had significantly wider eyes than in the baseline condition ($t(14)=2.11$, $p=0.027<0.05$, one-tailed, 95% CI [3.81% $+\infty$%], $d=0.55$). Figure 2.4 plots PSEs before and after meta-adaptation, again as percentage change relative to the inter-eye distance of a physically normal face (1.5 deg). The mean PSE increased by about 6%, comparable to the change measured in Experiment 1.

The meta-adaptation effect grew gradually over time. We divided the data into 4 bins based upon testing time, and Figure 2.5 plots estimates of PSE in successive bins (see Methods). A linear trend analysis revealed a significant upward slope ($t(13)=8.75$, $p<0.001$, one-tailed, 95% CI [+3.12 +∞]). We interpret this trend to indicate that repeatedly viewing the faces distorted by low-level adaptation caused PSEs to shift gradually towards wider eyes.

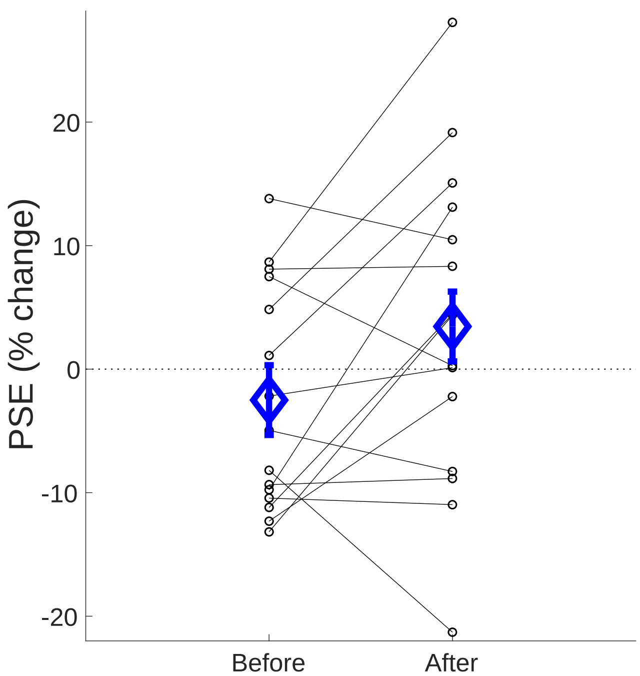

Figure 2.4 Results from Experiment 2.

Individual observers' percent change in PSE (relative to the eye distance of a normal face) are plotted with the group mean, both before and after adaptation. Error bars represent S.E.M difference score. Positive change corresponds to a shift of PSE to a face that has wider eyes than normal.

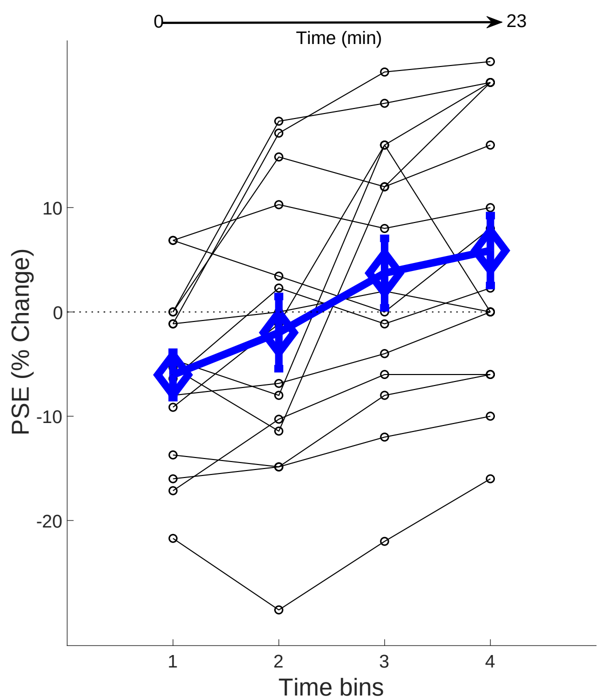

Figure 2.5 Time course results from Experiment 2.

Individual observers' PSEs and the group mean, computed in four time bins. Error bars represent ±1 S.E.M. The dotted line represents 0% change in PSE, i.e. the physical inter-eye distance of an undistorted face.

Discussion

Prolonged exposure to a normal face distorted by contrast adaptation produced robust changes in observers' face perception. That is, their visual systems adapted to distortions produced by adaptation. This "meta-adaptation" effect was reasonably large ($d > 0.5$ in two independent experiments) and grew gradually stronger over time. Changes in face perception were measured as changes in the face that appeared normal. These perceptual norm shifts were in the same direction as the distortions produced by the lower-level adaptation, leading the distorted face to appear more normal.

Our mechanistic interpretation of these results is that contrast adaptation affected early stages of visual processing; when a face was presented, signals modified by this lower-level adaptation were transmitted to later processing stages responsible for face perception, initially resulting in the percept of a distorted face. Over time, the later stages adapted to this altered input, shifting perceptual norms so that the percept of the face became more normal (Figure 2.1).

Our results and interpretation agree with previous reports showing that adaptation at early stages of visual processing can propagate downstream to later stages (Dhruv & Carandini, 2014; Patterson et al., 2014; Xu et al., 2008). Webster & Mollon (1995) suggested that later stages can themselves adapt to the altered input arriving from adapted earlier stages, in a way that is beneficial for color constancy.

Our results build on this past work, defining the concept of meta-adaptation, extending it to face perception, and showing that it can reduce illusory aftereffects produced by low-level adaptation, i.e. that it can mitigate the coding catastrophe. It is presumably beneficial because it allows earlier stages to take advantage of the coding efficiency gains that result from adaptation (Brenner, Bialek, & de Ruyter van Steveninck, 2000; Clifford et al., 2007; Wainwright, 1999; Wark, Lundstrom, & Fairhall, 2007) while reducing its potential negative consequences for later stages.

Note that the norm shifts we observed cannot be due to spread of the low-level adaptation to the location where we measured perceptual norms. The presence of low-level adaptation, which pushed the eyes outward, during norm measurement would cause faces with *narrower* eyes to appear normal. Our results were in the opposite direction. The meta-adaptation phenomenon we observed also may not hold for test faces with identities that differ in identity from the adapting face. This is likely because past work has established that our visual system maintains distinct perceptual norms for faces of different identities (Jaquet, Rhodes, & Hayward, 2008; Little, DeBruine, Jones, & Waitt, 2008a). Future work can test whether this is indeed the case.

Perceptual learning is another mechanism that may allow the visual system to overcome the coding catastrophe (Haak, Fast, Bao, Lee, & Engel, 2014; McGovern, Roach, & Webb, 2012).Past work has shown, for example, that training can allow observers to accurately judge moving stimuli, despite the presence of an ongoing motion aftereffect (McGovern et al., 2012). Learning may also allow observers to switch rapidly back and forth between adaptive states, which can minimize negative aftereffects of adaptation (Engel, Wilkins, Mand, Helwig, & Allen, 2016; Yehezkel, Sagi, Sterkin, Belkin, & Polat, 2010a). Perceptual learning, however, typically requires multiple days of experience and is often specific to the practiced task. Meta-adaptation may act more rapidly, generally, and automatically.

Probably the primary way the visual system copes with the coding catastrophe is to recover quickly from adaptation. Most demonstrations of adaptation's negative consequences rely on short-lived aftereffects that are visible once the adapting stimulus is removed. We have used such a paradigm here- the distorted face appeared only after the checkerboard pattern was removed. Aftereffects are generally used to isolate effects of adaptation from potential spatial interactions between the adapter and test, in the present case perceptual distortions that could have arisen if the checks and eyes had been presented simultaneously. Importantly, in natural viewing, the adapting stimulus may remain present continuously, requiring a solution such as meta-adaptation.

Meta-adaptation likely operates in domains other than faces. In principle, it could be used to aid perception of any feature that uses norm based encoding; these include color, blur, body perception, and others (Elliott et al., 2011; Rhodes et al., 2013; Webster & Leonard, 2008). Consider, for example, viewing an object in the presence of a high spatial frequency adapter (e.g. a house in the rain, though spatial interactions will be present in addition to adaptation). Adaptation should reduce the response to high frequencies, and make the object appear blurry. Meta-adaptation could in turn alter a high-level norm for blur, making the object appear more normal. Future work will determine how commonly vision, and other senses, use meta-adaptation to reduce effects of the coding catastrophe.

Chapter 3

Training on groups of similar faces strengthens group-level representations of trained faces

The visual system detects regularities within the perceptual environment and adjusts its processes to account for them. Examples of this adaptation include encoding and representation of light intensity (Shapley & Enroth-Cugell, 1984), contrast (Laughlin, 1981), color (Webster & Mollon, 1997) and textures (Olshausen & Field, 1996). Such a coding strategy may improve representational efficiency, for example by allowing neurons to use most of their limited dynamic range to encode the most common stimulus levels in the context (for a review on the functional benefits of adaptation, see Kohn, 2007).

Adjustment to environmental regularities may apply to higher level perception as well. In daily life, we often encounter groups of faces that may share similar facial features, for example people that belong to the same family, people that come from the same region, that are of similar age, or all of these. Could our visual system learn to optimize face representations based on regularities within faces that we frequently encounter?

One phenomenon, termed the 'other-race effect', provides some evidence that face representations can be influenced by the surrounding 'face environment'. In general, faces that belong to our own race are identified and recognized more easily and accurately than faces belong to other races; this own-race bias is found across various perceptual tasks, and cannot be solely attributed to differences in the face stimulus set (e.g. Byatt & Rhodes, 1998; Hayward, Favelle, Oxner, Chu, & Lam, 2017; for review, see Meissner & Brigham, 2001). The other-race effect may be due to face representations being optimized to fit our long-term experience viewing faces.

The other-race effect likely develops from long-term exposure to a face environment, but can our visual system switch between different representations on a faster time scale? For perception of some lower-level features like color (Engel et al., 2016; Li, Tregillus, Luo, & Engel, 2020) and blur perception (Yehezkel, Sagi, Sterkin, Belkin, & Polat, 2010b), our visual system appears to be able to 'remember' different perceptual environments and adjust more rapidly when they are re-experienced. For example, after 5 days of periodically wearing a pair of red lenses, observers' visual systems were able to immediately adjust upon donning the red lenses (Li et al., 2020). This almost instant switch between different processing states may allow observers to more efficiently adapt and adjust to commonly experienced environments.

Our perception of faces in daily life also involves interacting repeatedly with particular face environments: As mentioned earlier, we often encounter groups of faces that share similar features as we move through the environment. Therefore, it might be beneficial for our visual system to maintain multiple face representations and switch between them.

Here, we tested whether observers could learn in the medium term (days) to develop novel representations for faces present in a particular environment.

While prior work has shown robust perceptual learning in face processing, for example that training can reduce the other-race effect (e.g. Cavazos, Noyes, & O'Toole, 2019; Levin, 2000; Tanaka & Pierce, 2009), learning in these studies likely built upon long-term exposure to pre-existing face groups (e.g. own-race versus other-race faces). In the present study we tested whether observers could develop distinct representations for faces that belonged to the same racial identity but were seen in different contexts. Accordingly, we trained observers to identify individuals within two pre-defined groups of perceptually similar faces, within the same "race", where members of each group were encountered together, and at different times from the other group.

We used a visual search training paradigm to establish face environments and encourage observers to establish distinct representations for the groups of faces. Visual search is commonly used as a perceptual training paradigm to improve recognition of objects (Sigman & Gilbert, 2000) and faces (Levin, 2000). Our observers were trained to locate a target face among distractor faces that belonged to the same pre-defined group. Observers trained on different groups of faces in separate sessions. Since difficultly of a visual search task partly depends on the target's similarity to distractors (Treisman & Gelade, 1980), improvement in visual search may be attributed to the target appearing more and more distinct perceptually from the distractors. We hypothesized that alternating visual search sessions in which observers learned to tell apart faces that came different groups would help observers establish distinct representations for those groups of faces.

To visualize how groups of faces were perceived, and whether groupings of faces were strengthened by training, we asked observers to rate how similar each face encountered in our experiment was to each other one. We hypothesized that as groupings of faces strengthens, face pairs between different groups should appear more dissimilar than face pairs belonging to the same group.

To further show how training affects group-level representations for trained faces, we conducted a hierarchical clustering analysis on their responses. Clustering analyses can be performed by many unsupervised machine learning algorithms that join similar objects (or perceptually similar faces in the present study) into groups.

In sum, in this study observers were asked to rate perceptual similarity pairs of faces, before and after three consecutive days of training recognizing members of two groups of faces presented separately. We reasoned that if learning could establish distinct representations for the trained groups of faces, then after 3 days of learning 1) trained faces would appear less similar to each other, indicating that visual search training was effective; 2) evidence for face grouping would strengthen: Faces between different groups would appear more dissimilar than faces that came from the same group. 3) clustering of the trained groups of faces would strengthen after training.

Methods

Participants

16 observers participated in this experiment. Sample size was determined from prior studies on perceptual face space (e.g. Byatt & Rhodes, 2004) and

perceptual learning of facial identities (e.g. Apps & Tsakiris, 2013), as well as an unpublished pilot version of this experiment. All experimental procedures were approved by the University of Minnesota Institutional Review Board (IRB 1006M84457) and conformed to the principles of the Declaration of Helsinki.

Stimuli

Database

Faces in this experiment were selected from the Chicago Face Database (CFD; Ma, Correll, & Wittenbrink, 2015). The CFD is a free face database containing 158 standardized photographs of faces, all taken in a consistent setting and manner. These faces are of different race, gender, and age. Additionally, the CFD provides various objective physical measurements of the features of each face, for example median luminance of the skin tone, nose width, lip thickness, cheekbone height, etc.

Selecting Faces

We first conducted a principal component analysis (PCA) to aid selection of subsets of faces suitable for this experiment. We included all objective measurements provided by the CFD (39 total) in our principal component analysis (as input), and given that the first 6 principal component explained more than 95% of the variation in the objective measurements, we used these principal components (and principal component scores) to construct a six-dimensional ‘physical representational space’ of the faces in the database. Within this space, faces that were close together shared similar physical features, compared to faces that were further apart.

Figure 3.1 shows faces that we selected from the CFD, and pairwise distances of all face pairs in the physical face space. Our main goal of face selection was to define 2 groups of same-race faces (to avoid pre-existing groupings) that initially appear similar to naïve observers, and another group of faces that belonged to a different race as a control. We selected 3 groups of faces, with 5 faces per group, using Euclidean distance in the PCA physical face space to quantify how ‘close’ or ‘far apart’ any given face pair was. To pick the faces we identified groups of faces such that:

1. all faces were the same gender.
2. faces in 2 groups were labeled in the database as the same race. These were the groups that were trained. Faces in the third, untrained, group was from a different race.
3. the faces within a group were relatively close to each other, compared to the average distance between faces in the PCA space.
4. the faces within a group were closer to each other than faces from different groups.
5. the faces in the trained groups were closer to each other than to faces in the untrained group.

We labeled the two groups that were close to each other in the physical face space as trained group 1 and trained group 2. We referred to group 3 as the ‘untrained group’.

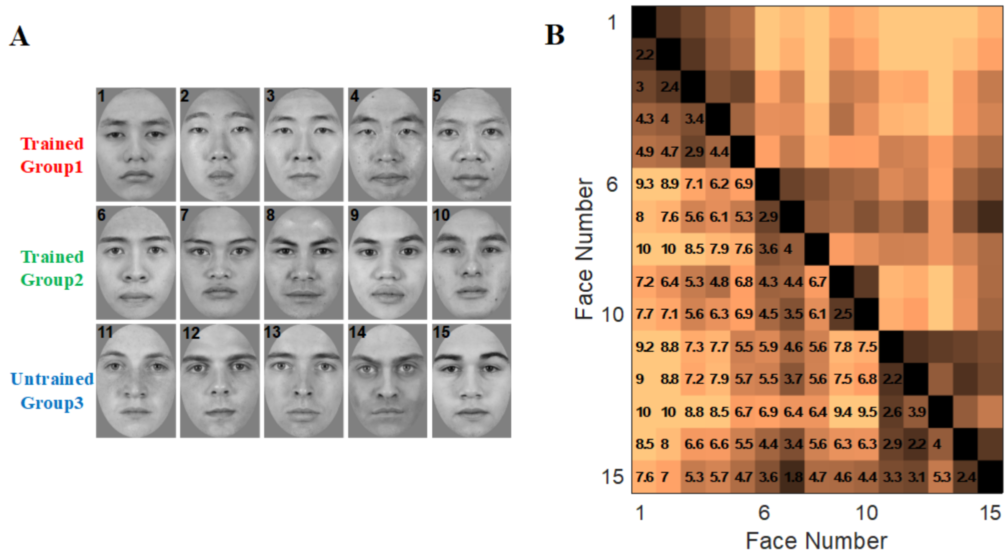

Figure 3.1 A: Faces used in this experiment. These faces were selected from the CFD. Number on each face corresponds to "Face Number" on the pairwise distance matrix. **B**: Pairwise distance matrix of all 15 faces selected. Distances were measured as the Euclidean distance of face pairs on the 6-dimensional physical face space. Groups are visible as rectangles of high similarity (low distances) along the diagonal.

Face pre-processing

We believed that a relatively standardized and homogeneous set of faces would better facilitate the formation of group representations during the training in visual search. Selected faces were resized to 4 X 5.8 deg visual angle and presented through an oval shaped aperture. Some faces that we selected contained visible hairline or hairstyle within this window, and we removed hairstyle from these faces using Adobe Photoshop. Faces were presented on a 20 X 18 deg uniform grey rectangular background with a mean luminance of 37 cd/m^2. Mean RMS contrast of all faces was 0.12. We added noise to our face images to encourage a more holistic viewing of faces: on each trial, additive pixel-wise gaussian noise was drawn from a normal distribution with a mean of 0 and variance of 3dB.

Experimental Procedure

Overview

In this experiment, observers first completed two sessions of a similarity rating task, where they judged all possible pairs of the 15 faces. Observers next trained in identification of members of the groups using a visual search task: Target and distractors faces were chosen from the same group, and the two nearby groups were trained in alternate sessions. After three consecutive days of training, observers completed another session of the rating task. Each observer took 5 days to complete the experiment.

Similarity Rating Task

Fig 3.2 shows a sample trial of the rating task. Observers rated, on a scale of 1-5, how similar any given pair of faces appeared to them. Specifically, observers were instructed to rate how likely were pairs of faces related to each other ('how likely are these faces siblings?'), with '1' being the most likely (hence face pairs appeared more similar) and '5' being the least likely. We believed this is a more intuitive and conceptual way to measure perceptual similarity, and further encourages holistic face perception. Each observer viewed and rated 105 face pairs in total (all pairwise combinations of 15 faces).

Observers sat 65cm away from the computer screen, and viewed the 4 X 2.8 deg faces (Fig 3.1). Pairs of faces always appeared one to the left, and the other to right hand side of the fixation marker. Exact location of the face pairs was slightly jittered on each trial, moving up or down by approximately 3.4 deg, and left or right by 0.4 deg. This was done to minimize retinotopic afterimages.

Each trial started with 1 sec of blank screen (20 by 18 deg uniform grey background with additive Gaussian noise), followed by a 2 sec presentation of a face pair. Observers were instructed to fixate a central fixation marker during face presentation, after which the faces were replaced with a mean gray screen for 2 sec. Observers then responded, using the number keys 1-5 to indicate how similar each pair of faces appeared. After 2 sec, a new pair of faces appeared, and this continued until observers completed rating the full 105 pairs of faces. One full run of the rating task took approximately 12 minutes to complete.

Each observer completed two runs of this rating task, before participating in 3 consecutive days of the search task. Upon completion of the search task, observers again completed a single run of the rating task.

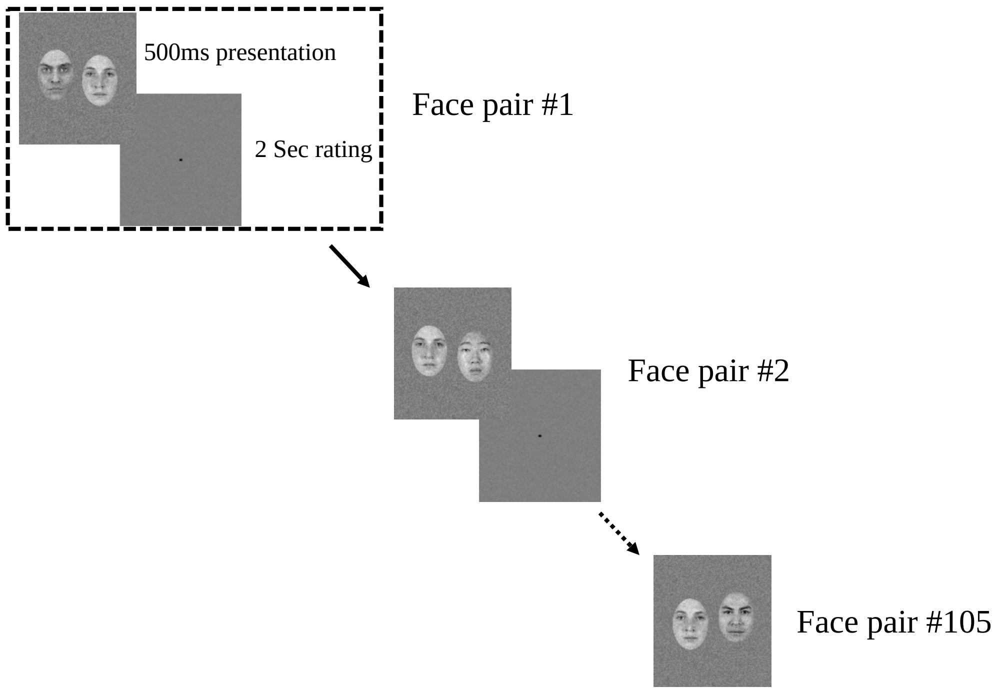

Figure 3.2 Similarity rating task. On each trial, observers were shown a randomly selected pair of faces for 500 ms, and had 2 seconds to respond, on a scale of 1-5, how likely they believed the face pair to be siblings.

Visual search task

Figure 3.3 shows a sample trial of the search task. Observers were instructed to locate a target face, among 15 other "distractor" faces. Target and distractor faces were shown on a 30 by 24 deg uniform grey background with a mean luminance of 37 cd/m^2, and each face was 4 by 2.8 deg visual angle. All 16 faces were shown in a 4 by 4 grid: we divided the 30 by 24 uniform grey background evenly into 16 rectangles that formed this 4 by 4 grid. Faces were placed at the center of each rectangle, but randomly jittered by 0 deg to 0.4 deg in both X and Y directions. This jittering was added to account for any possible global pattern effects that may have helped searching for the target without processing each face independently.

The target face was always shown to observers at the center of the screen before every trial. Observers were instructed to locate the target face as quickly and as accurately as possible.

When observers were ready to begin a trial, they pressed any key on the keyboard and the array of faces appeared. Upon locating the target face, observers first pressed the space to allow recording of reaction time, and then left clicked on the face they believed to be the target face. Observers were given feedback after each trial: If the selected face was indeed the target, a green rectangle would appear around the selected face, at the same time a high pitch (700Hz) beep played; otherwise, a red rectangle would appear around the actual target face, and a lower pitch (200Hz) beep played. Observers were instructed to skip a trial by right clicking anywhere on the screen if they forgot the target face.

Each visual search session consisted of 10 individual blocks, and each block had 10 trials. Within each block observers always searched for the same target face, and after completing the 10 trials observers a new target face was presented.

During each search session, target and distractors all came from the same group of faces, either from trained group 1 or trained group 2, and each session took approximately 25 minutes to complete. Observers searched for faces from different groups in alternate sessions. Each observer completed the search task on

3 consecutive days, with two sessions per day, one in the morning and another in the afternoon.

A

B

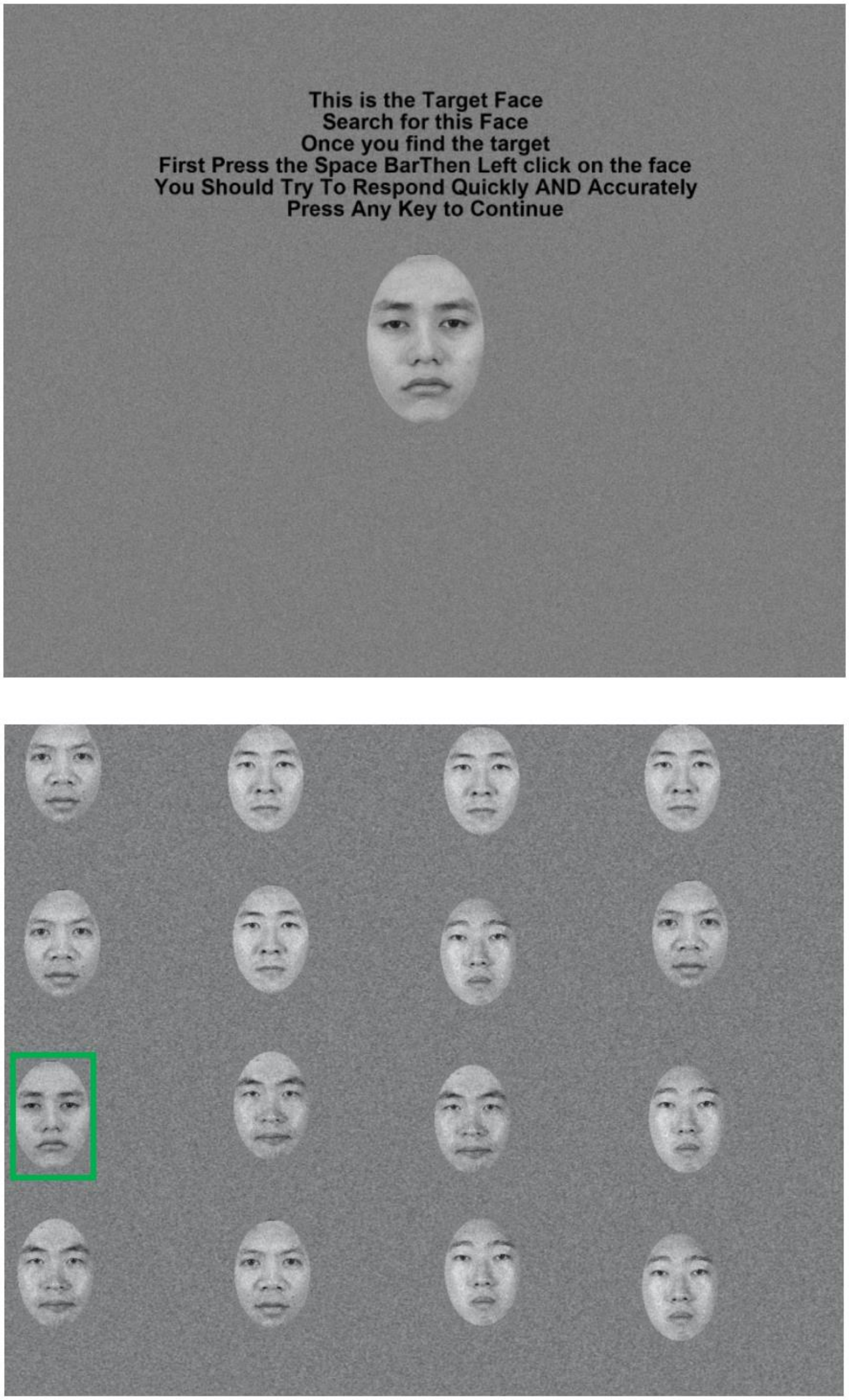

Figure 3.3 Visual search task. A: Observers were shown a target face and were instructed to locate this face as quickly and as accurately as possible. B: The 16 by 16 search grid. Upon locating the target, observers first pressed the space bar and clicked on the target. If observers correctly located the target, a green rectangle was displayed; otherwise, a red rectangle surrounding the correct target location appeared.

Results

Visual search task

Figure 3.4 shows the average of the median search time for each observer (n=16) over the course of the 3 training days. Observers took significantly less time to locate the target face across days (average search time between the first and the last session decreased by 1.02 seconds, $t(15) = 8.19$, $p<0.001$), and the decreasing trend is easily visible.

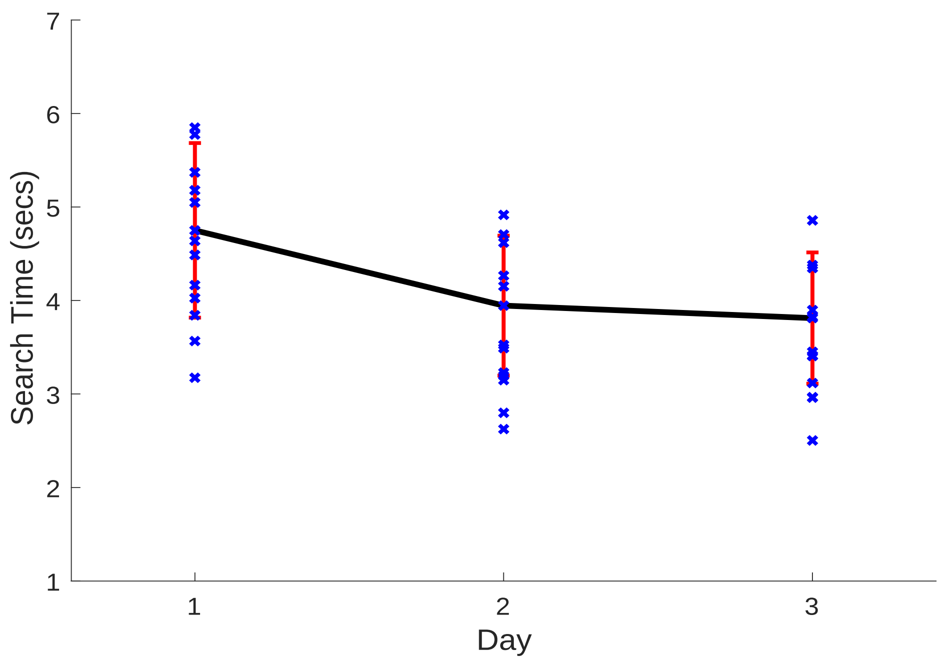

Figure 3.4 Average search time across three days of learning. The black line plots the average (n=16) of the median reaction time. Blue crosses show individual median reaction times. Error bars represent S.E.M between observers.

Similarity rating task

<u>Before training</u>

Figure 3.5A shows the average dissimilarity scores for all pairs of faces (105 pairs). A high dissimilarity score indicates relatively low similarity between

the two faces. Thus, the dark regions mark pairs faces that appear relatively similar (e.g. most pairs of faces 11-15) and the light regions indicate groups of faces that appear relatively dissimilar (e.g. pairs with one face from 11-15 and another face from 1-10).

We were interested in whether similarity ratings provided evidence that participants were sensitive to the structure of the three predefined groups of faces, even before training. Recall that trained groups 1 and 2 were of the same race, and were more physically similar to each other, than to the untrained group 3, which was of a different race. In addition, members of each group were more physically similar to each other than to members of other groups.

Observers appeared to be sensitive to physical differences that defined the groups. Figure 3.6B plots the average rating for "**within group**" face pairs (where both members of the pair came from the same predefined group, faces 1-5, 6-10, and 11-15 for trained groups 1, 2, and untrained group 3 respectively, see Figure 3.1A) and "**between group**" face pairs (the two faces in the pair were members of different groups).

Members of a group appeared more similar to each other than to members of other groups. Specifically, observers' dissimilarity ratings for within groups pairs were lower than for between group pairs (difference in dissimilarity = 1.15, $t(15)=12.04$, $p<0.001$) (Compare left and right sets of bars in Figure 3.5B).

We refer to this difference (between group-within group) as the 'grouping index', which serves as a metric for grouping strength: A larger difference indicates between-group pairs appeared more dissimilar than within-group pairs.

Throughout the rest of the analyses, we will use the grouping index to assess grouping strength.

Observers also appeared to be sensitive to the greater physical similarity between the two trained groups relative to the untrained group. Ratings for between group face pairs differed significantly between the three groups ($F(2,30)=18.07$, $p<0.001$), and pairwise comparisons showed that dissimilarity between face pairs where one face was from each of the two trained groups was reliably less than dissimilarity between pairs where one face was from the trained groups and one from the untrained group (mean difference = 1.06, $t(15)=5.72$, $p<0.001$) (compare first bar on right of Fig 3.5B to adjacent two).

Finally, the three groups did not differ greatly in terms of the perceived similarity within the group. Ratings of dissimilarity for face pairs within the predefined groups did not differ significantly across groups ($F(2,30)=2.47$, $p>0.05$). However, a post-hoc comparison revealed that, face pairs within the untrained group appeared more slightly more dissimilar than face pairs within the trained groups ($t(15)=2.09$, $p<0.05$).

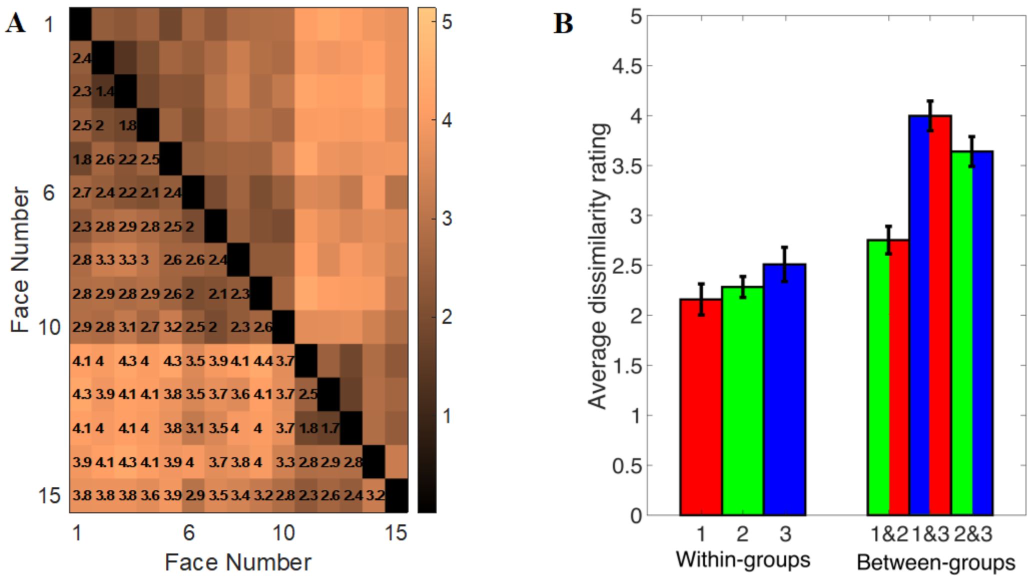

Figure 3.5 Dissimilarity ratings for face pairs before training. A: Average dissimilarity rating matrix for all face pairs.
B: Average rating scores for face pairs that came from the same predefined groups ("Within-groups") and for face pairs that came from different groups ("Between-groups"). Higher rating score corresponds to face pairs appearing more dissimilar. X-axis labels, '1','2' and '3', correspond to face pairs within trained groups 1, 2 and untrained group 3.

<u>After training</u>

We hypothesized that training might strengthen evidence for the pre-defined groups. Figure 3.6A shows the dissimilarity ratings for all pairs of faces after training using the visual search task, and Figure 3.6B shows average perceptual dissimilarity ratings for within group and between group face pairs. Observers again appeared to be sensitive to the three groups, with a grouping index reliably above zero (mean difference= 1.10, $t(15) = 9.62$, $p<0.001$); compare left and right sets of bars in Figure 3.6B).

Even after training, however, the two trained groups remained more perceptually similar to each other than to the untrained group. Dissimilarity ratings for between group face pairs were lower for the two trained groups (i.e. both faces were the same race) vs when one of the two was untrained. Statistically, between group face-pair dissimilarity differed significantly across the compared groups ($F(2,30)=34.32$, $p<0.001$), and pairwise comparisons showed that dissimilarity between pairs where one member was from each of the two trained groups was less than pairs where one member was in a trained group and one was in the untrained group (mean difference in dissimilarity = 0.78, $t(15)=6.12$, $p<0.001$).

Finally, ratings for within group face pairs now differed significantly across groups ($F(2,30)=4.38$, $p=0.02<0.05$). Trained group 2 had reliably higher within group dissimilarity than the other two groups (1 vs 2=0.41, $p=0.03<0.05$; 2 vs untrained group 3 =0.49, $p=0.02<0.05$; 1 vs untrained not significant). Some reasons for this will be discussed following clustering analysis, below.

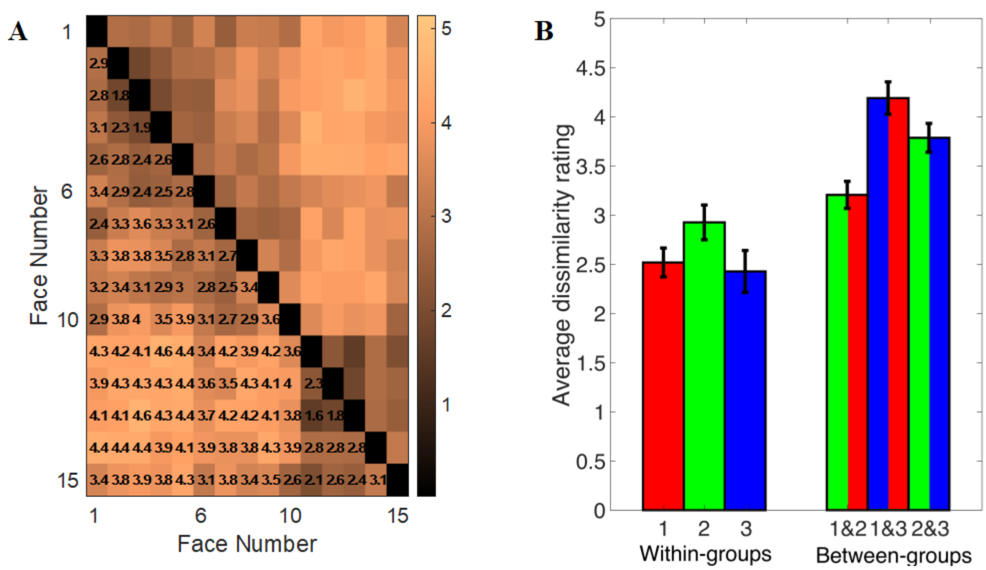

Figure 3.6 Dissimilarity ratings for face pairs after training. A: similarity rating matrix after training. B: Average rating scores for face pairs that came from the same groups ("within-groups") and for face pairs that came from separate groups ("Between-groups"). Higher rating score corresponds to face pairs appearing more **dissimilar**. X-axis labels, '1','2' and '3', correspond to face pairs within trained groups 1, 2 and untrained group 3.

Comparing dissimilarity ratings before and after training

Comparing similarity ratings before and after training on the search task allowed us to test more directly whether learning affected representations of the trained groups. As expected, participants learned to better distinguish faces within the trained groups (Figure 3.7). Training increased dissimilarity for pairs within the trained groups by an average of 0.59 in group 1 ($t(15)=2.76, p=0.01<0.05$), and of 1.06 in group 2 ($t(15)=3.53$ $p<0.005$). Face pairs within the untrained group however did not change following training (mean difference in dissimilarity score=0.19, $t(15)=0.59$, $p>0.05$). The increase in dissimilarity for pairs within the trained groups was significantly greater than pairs within the untrained group (mean difference = 0.36, $t(15) = 3.32$, $p=0.02<0.05$).

Observers also learned to better differentiate the two trained groups, as dissimilarity between the two trained groups increased following training (face pairs with one member from group 1 and one from group 2 increasing in dissimilarity by 0.45, $t(15)=4.20$, $p<0.001$). Dissimilarity between the trained and untrained groups increased relatively little; for pairs where one face was from trained group and one from an untrained group, dissimilarity increased by 0.04, ($t(15)=0.42$, $p>0.05$). This increase in dissimilarity was reliably less than the increase in face pairs between the two trained groups (mean difference=0.41, $t(15) = -2.25$, $p=0.02<0.05$).

Above we defined overall evidence for the three predefined groups as a positive grouping index. i.e. when between-group pairs overall appeared significantly more dissimilar than within-group pairs. We hypothesized that training would increase this index for the three pre-defined groups. However, it did not change much after training compared to before (mean difference in grouping index = 0.04, $p>0.05$).

In sum, comparing similarity ratings before and after training showed that: 1) face pairs within the trained groups appeared less similar following learning; 2) face pairs between trained groups ('1&2' in Figure 3.7) also appeared less similar following learning; 3) similarity within the untrained group did not change significantly; also, perceptual dissimilarity in face pairs between trained and untrained groups ('1&3' and '2&3' in Figure 3.7) changed relatively little. 4) evidence for pre-defined groups as measured by the grouping index was strong, both before and after 3 days of training. Training did not change the grouping index. Together, these results provide some evidence that learning strengthened representations of the trained groups.

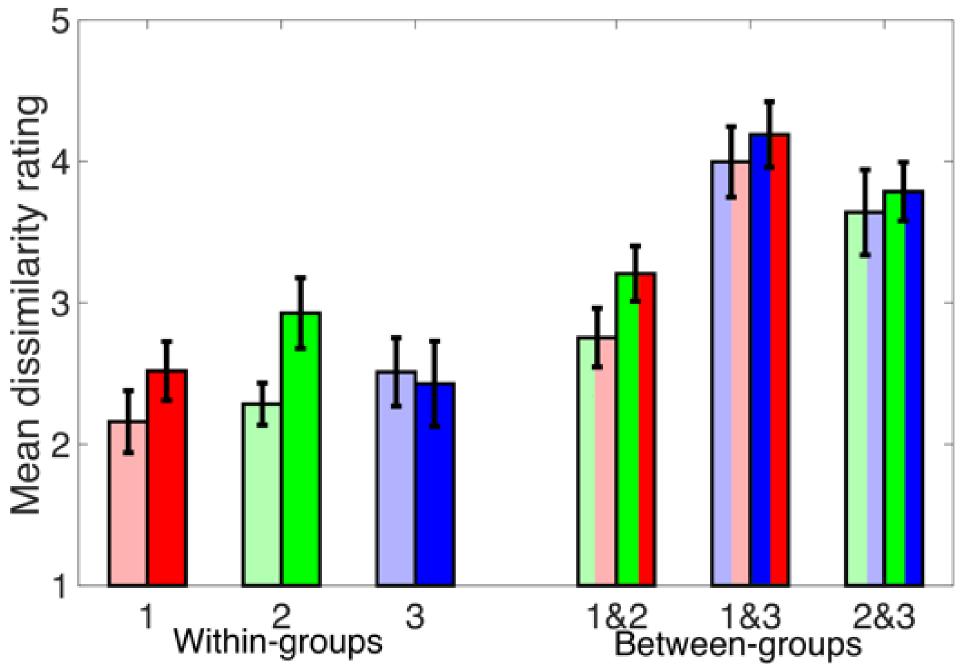

Figure 3.7. Average dissimilarity ratings for within and between-group face pairs, before and after learning. Bar colors represent the 3 pre-defined groups; saturation of the colors represents average similarity rating before (less saturated) and after (more saturated) training.

Hierarchical clustering analysis of dissimilarity ratings

To more directly examine representations of groups of faces, we used exploratory hierarchical agglomerative cluster analysis. The analysis begins by assigning each face to its own cluster and then merges clusters to form new ones, such that at each step the merging operation forms the new cluster with the smallest average dissimilarity across all pairs in the cluster.

Figure 3.8 visualizes the hierarchical clustering results on a dendrogram. The X-axis of the dendrogram shows faces that were collected into different clusters, with pairs of faces belonging to the same cluster arranged next to each other. The y-axis of the dendrogram plots the measure of cluster strength used to merge clusters- the average perceptual dissimilarity of all face pairs within clusters- with lower numbers indicating stronger clusters. Moving down the dendrogram yields more clusters that each contain fewer faces, but with greater similarity (less dissimilarity) within the cluster.

Examining the dendrogram based on similarities measured before training (Fig 3.8A), starting from the top, shows clear evidence for two clusters, with the trained groups' faces 1-10 and the untrained faces falling into distinct clusters. As we move further down the dendrogram, the large cluster of faces divides in two, forming three clusters that reflected the predefined groups: Faces 1-5, faces 6-10, and faces 11,12,13,15, but with face 14 considered its own cluster. Moving further down the tree, these three clusters rapidly divide into many small clusters of size 2.

Figure 3.8B shows the dendrogram after training: Again, starting from the top, there is evidence for two clusters comprising the trained and untrained faces; however, this grouping relatively quickly breaks into three groups, with faces 11-15, faces 1,7,8,10, and faces 2,3,4,5,6,9 forming 3 distinct clusters. Moving further down the dendrogram, these 3 clusters again further divide into smaller clusters.

Interestingly, some faces that formed tight clusters after 3 days of training were members of different pre-defined groups: For example, face 1 from trained group 1 joined a cluster with faces 7,8 and 10 from trained group 2; on the other hand, faces 6 and 9 joined another cluster with other members of pre-defined trained group 1. These results could explain the observed larger increase in perceptual dissimilarity for pairs within trained group 2, compared to pairs within other pre-defined groups, given more faces from pre-defined group 2 joined a cluster with faces from pre-defined group . These shifts provide further evidence that training led observers to form different representations of the groups of faces.

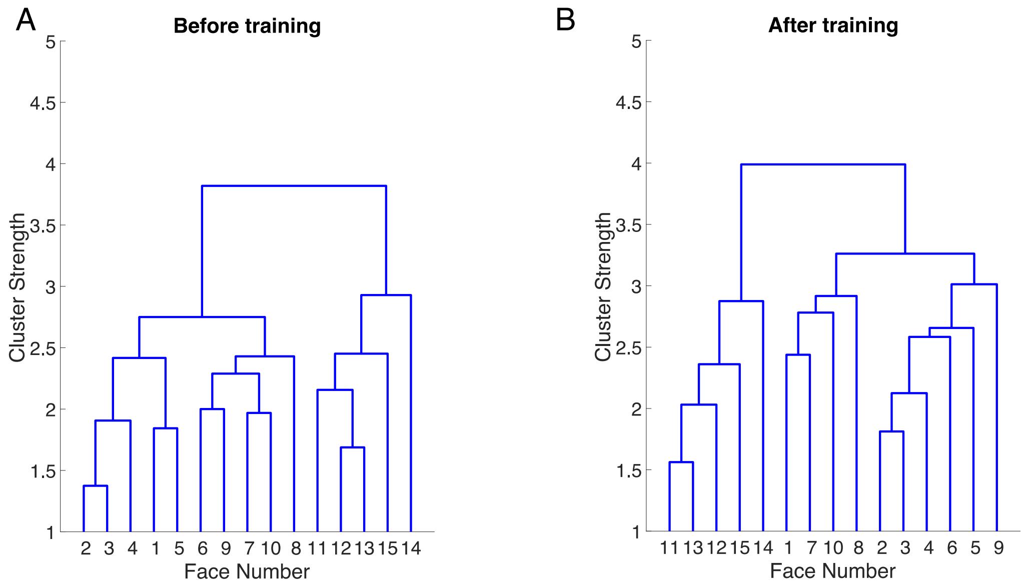

Figure 3.8. Dendrograms of before and after training dissimilarity ratings. Faces were arranged based on pairwise similarity on the x-axis, and the y-axis plots cluster strength, which measures the average perceptual dissimilarity of all faces within clusters. Smaller values correspond to tighter clusters.

Next, to formally evaluate the evidence for different clusterings shown in Figure 3.8, we computed our standard metric, grouping index, for the two- and three-cluster solutions identified in both data sets (Figure 3.9). A larger grouping index indicates a stronger clustering of faces.

Prior to training, the two-clusters solution produced a slightly higher grouping index than the 3-cluster solution (difference in grouping index (two-cluster vs three cluster) = 0.19, $t(15)=0.53$, $p>0.05$). After learning, the three-cluster solution yielded a larger grouping index (difference = 0.21, $t(15)=2.60$, $p<0.01$). The interaction between learning and clustering solutions is statistically significant (p=0.02<0.05). Thus, prior to training there is greater evidence for a 2 group than three group representation, while the reverse was true following training.

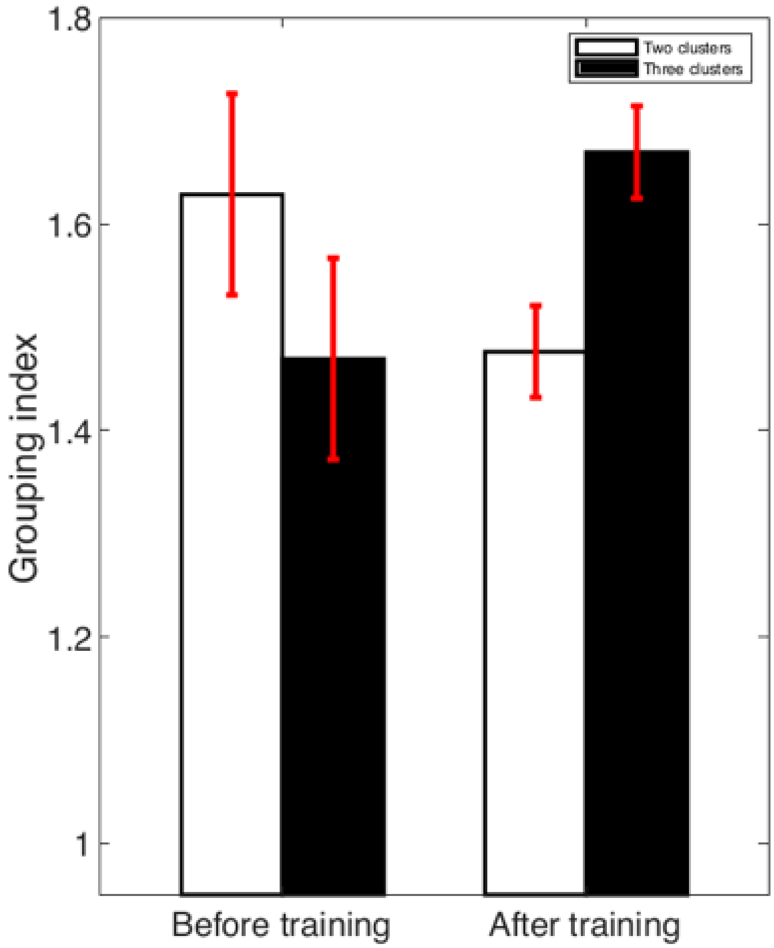

Figure 3.9 Grouping index for two- and three-cluster solutions, before and after 3 days of learning. Error bars represent S.E.Ms of the difference in perceptual dissimilarity for n=16 observers, for different clustering solutions before and after training.

Figure 3.10 shows the perceptual dissimilarity matrices rearranged to reflect the stronger of the 2- or 3-cluster solutions for each data set. In Figure 3.10A (before training dissimilarity matrices), the two pre-defined trained groups of faces (faces 1-10) were tightly clustered, and untrained faces (11-15) formed a different tight cluster. Following training (Figure 3.10B) two distinct clusters formed among trained faces: Faces 1,7,10 and 8 were perceptually more similar than other faces, as well as faces 2,3,4,5,6, and 9.

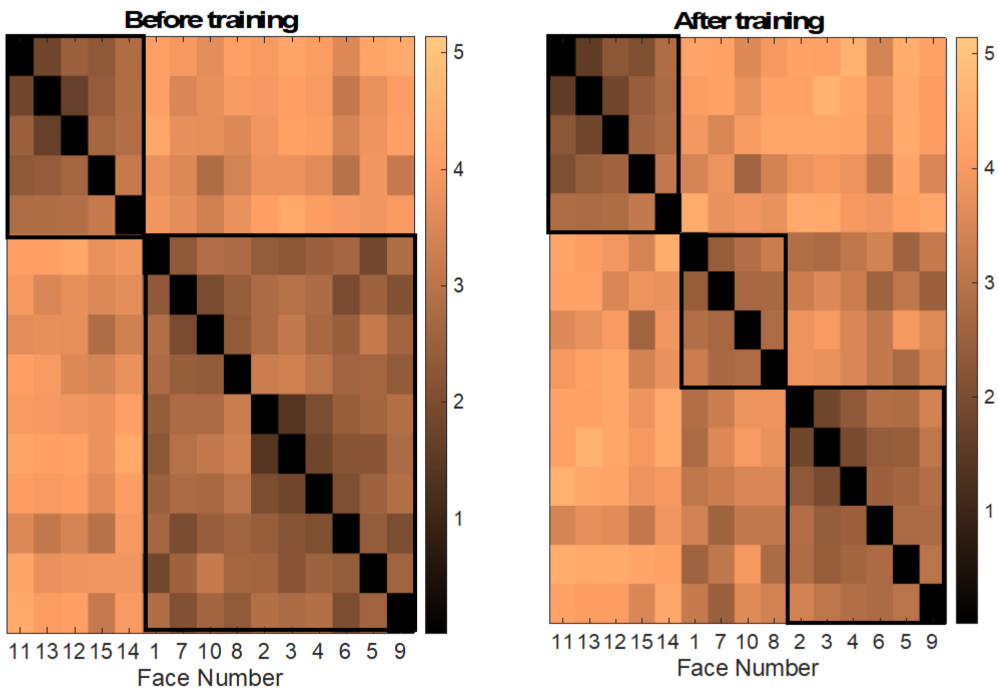

Figure 3.10. Dissimilarity matrices, with clusters of faces enclosed with rectangles. X-axes of the matrices were rearranged so that faces belonged to the same cluster after training were close to each other on the X-axes.

Discussion

Our results showed that after training designed to strengthen representations of two groups of faces that initially appeared similar, observers developed novel group-level representations for trained faces. Evidence for this conclusion includes: 1) Faces that belonged to the trained groups became more dissimilar to each other, both for face pairs that came from the same trained group and face pairs that came from different trained groups. 2) Perceptual dissimilarity for face pairs that contained untrained did not greatly change, indicating that learning was specific to the trained groups. 3) Although, training did not increase a grouping index for our pre-defined groups, clustering analysis showed that training produced new clusters of faces among trained faces, with evidence for two distinct clusters amongst the trained faces being greater following training.

Our results agree with previous studies showing that perceptual learning can improve performance on face perception tasks: e.g. recognition (Apps & Tsakiris, 2013) and discrimination (Yang et al., 2020). The present study's results are consistent with different models of perceptual learning: For example, there could be changes in tuning of groups of neurons encode trained faces, where neural tuning to trained faces might sharpen with practice (e.g. Schoups, Vogels, Qian, & Orban, 2001); alternatively, groups of neurons best tuned to search for the target among distractors might become more active through a selective reweighting process that maximizes outputs of the groups of neurons 'contribute' most to the search task (e.g. Dosher, Jeter, Liu, & Lu, 2013).

We also showed that observers could learn to form more distinct representations for **groups** of faces. This novel finding adds to current literature on perceptual learning and the other-race effect by showing that our visual system is capable of learning to represent groups of faces that appear more as one homogeneous group to naïve observers. Importantly, this finding is highly relevant to everyday face perception, as we frequently encounter groups of faces that share some facial features (e.g. come from the same race group) but are also different in other fine-grained facial features that could potentially put faces that belong to the same coarse category into sub-groups.

In related work, a few machine learning studies have investigated how faces belong to the same general group but come from different perceptual environments are processed: for example training different machine learning algorithms to classify Chinese/Koreans/Japanese faces (Wang, Feng, Liao, Luo, & Xu, 2018), to classify Chinese sub-ethnicities (Duan et al., 2010), Myanmar faces (Htake, Tin, & Sein, 2011) and Indian faces (Katti & Arun, 2019). Algorithms trained with certain facial features (mouths and facial contours) performed reliably better at the classification task than trained with other facial components (e.g. Katti & Arun, 2019). Future studies could uncover the learned facial features useful for performing such classification tasks.

In sum, our results suggest that visual search training helped observers strengthen representations for groups trained faces. To provide stronger evidence for this hypothesis, we propose a more direct behavioral measurement: Face adaptation aftereffects (FAE). As discussed in Chapter 1, these effects are specific to properties of the adapting faces. Transfer of the FAE (or lack thereof) can serve as an index of whether faces share similar encoding and representation. In Chapter 4, we measured transfer of FAE before and after 3 days of learning to test for evidence of group representations.

Chapter 4

Face adaptation reveals dissociable representations for groups of faces after perceptual learning

In chapter 3 we showed that after 3 days of training, face pairs within a trained group as well as face pairs between different trained groups appeared more dissimilar, while untrained faces did not change their appearance. Clustering analysis further revealed that training formed distinct new clusters of faces among trained faces. Our interpretation was that following our training paradigm, observers were able to form distinct perceptual representations for different groups of faces, which resulted in these faces appearing more dissimilar. ‘

Here we more directly test a prediction of these results. If our visual system established more separate and specific representations for groups of trained faces, then we should be able to observe an increase in representational/coding specificity as measured by the “psychophysicists electrode”, adaptation. In chapter 1, I explained how visual adaptation allows psychophysicists to infer neural mechanisms from behavioral measurements: In short, adaptation effects are generally selective, meaning that neural modulation caused by adaptation are specific to the neurons that respond to the adapting stimulus (e.g. summarized in Kohn, 2007). Measuring transfer of face adaptation effects across faces should allow us to infer whether faces from different pre-defined groups are encoded by the same or dissociable mechanisms.

In this chapter, we will measure face adaptation aftereffects when observers are exposed to one distorted face and then ask whether a different face also appears distorted. If adapting to one distorted face (the “adaptor”) affects subsequent perception of another face (the “test”) by a relatively large amount, then we can infer more strongly shared coding mechanisms between the adaptor and the test. If face adaptation effects transfer relatively little, then we can infer more independent encoding. Previous studies have used this logic to identify when different faces are encoded and represented using different mechanisms. For example, face adaptation aftereffects were found to be contingent upon gender (Little, Debruine, & Jones, 2002) age (Barrett & O’Toole, 2009), species and facial expression(Little et al., 2008a).

In this experiment, we used face adaptation to test whether multiple days of training changed how the trained faces were represented. Our observers (n=10) first completed a face adaptation paradigm, in which we measured transfer of FAE across the three groups of faces that we selected and defined in experiment 2. Observers then completed 3 days of the visual search task, to help them establish separate representations for the trained groups of faces. After 3 days of training, observers again completed the face adaptation paradigm. Additionally, observers completed the similarity rating task used in experiment 2, at both timepoints.

We compared transfer of FAE across groups of faces, before and after training. We expected training to produce more distinct representations of the two trained groups, and so expected to observe a decrease in transfer of adaptation between them.

Methods

Participants

10 observers (5 Asian, 5 Caucasian) participated in this experiment. Sample size was determined from previous studies on face adaptation aftereffects (e.g. Webster, Kaping, Mizukami, & Duhamel, 2004), perceptual learning of facial features (Apps & Tsakiris, 2013), as well as an unpublished pilot experiment. All experimental procedures were approved by the University of Minnesota Institutional Review Board (IRB 1006M84457) and conformed to the principles of the Declaration of Helsinki.

Figure 4.1 Faces used in Experiment 3. Each row is one group of faces. The first two groups/rows were used in the search task, as observers were trained to tell them apart. The 'test face' used in the adaptation task is enclosed by a black border, and the three adaptor faces are enclosed by colored borders.

Stimuli

Face database, selection, and pre-processing

Figure 4.1 shows all faces used in this experiment; these faces were the same those used in Chapter 3. Selected faces were resized to 4 X 5.8 deg visual angle, and presented through an oval shaped aperture. For faces that had hair visible through the aperture we removed hair using Photoshop to make faces more homogenous.

Selecting the adaptor and test faces for the adaptation task

Our goal was to select one adaptor face from each pre-defined group, and one test face overall so that we could measure patterns of transfer of face adaptation effects when observers adapted to different adaptor faces and were subsequently shown the test face. Specifically, we wanted to see if transfer of face adaptation effects between faces that belonged to different trained groups changed after training. Additionally, we measured transfer of adaptation between faces that belonged to the same trained group, and faces that belonged to trained and the untrained group as controls.

We conducted a pilot similarity rating experiment (n=8; results were pooled in Chapter 3 similarity rating data) and used the pairwise dissimilarity rating data to select the appropriate adaptor and test faces. Figure 1 shows the adaptor and test faces selected: Adaptor faces from the 3 pre-defined groups are enclosed by colored borders, and the test face (from trained group 1) is enclosed by a black border. For adaptors from the trained groups, we selected the face in each group that appeared, on average, most similar to other members within the same group. For the untrained group adaptor, which served as a control to see if face adaptation effects occurred for all faces, we selected the face that appeared least similar to the trained groups. Lastly, for the test face we selected a face that had high similarity to faces in trained group 1.

Experimental Procedure

Overview of the experiment

Figure 4.2 shows an overview of the 3 week testing period. Observers completed the face similarity rating task before and after training on faces with visual search, following almost identical procedures to Chapter 3. We added an adaptation task before and after training, to reveal potential changes in representation of the faces. During the first week, each observer first completed the adaptation task, then the similarity rating task. In week 2, each observer spent 3 days (2 daily sessions) training with the faces using the same visual search paradigm. Observers then again completed the similarity rating task. In the third week of the experiment observers repeated the adaptation task.

Week 1: Pre-training Adaptation and Rating	**Week 2: Training and rating**	**Week 3: Post training adaptation**
Day 1: obtain IRB consent; practice the adaptation task	**Day 1-3**: Search task, two sessions per day	**Day 1-3**: Adaptation task: two sessions per day
Day 2: practice the adaptation task	**Day 4**: Rating task	
Day 3: Adaptation task: two sessions		
Day 4: Adaptation task: two sessions		
Day 5: Adaptation and rating task		

Figure 4.2 Overview of the 3-week experiment.

Figure 4.3 shows the adaptor faces selected from the 3 pre-defined groups. During each adaptation session, observers completed two blocks of a normality judgment task (described below). In the first, observers viewed an undistorted version of the adaptor, as the 'baseline' block. In the second, 'adaptation' block, a distorted adaptor as shown (fig 4). Observers viewed the adaptors from the three different face groups in separate sessions.

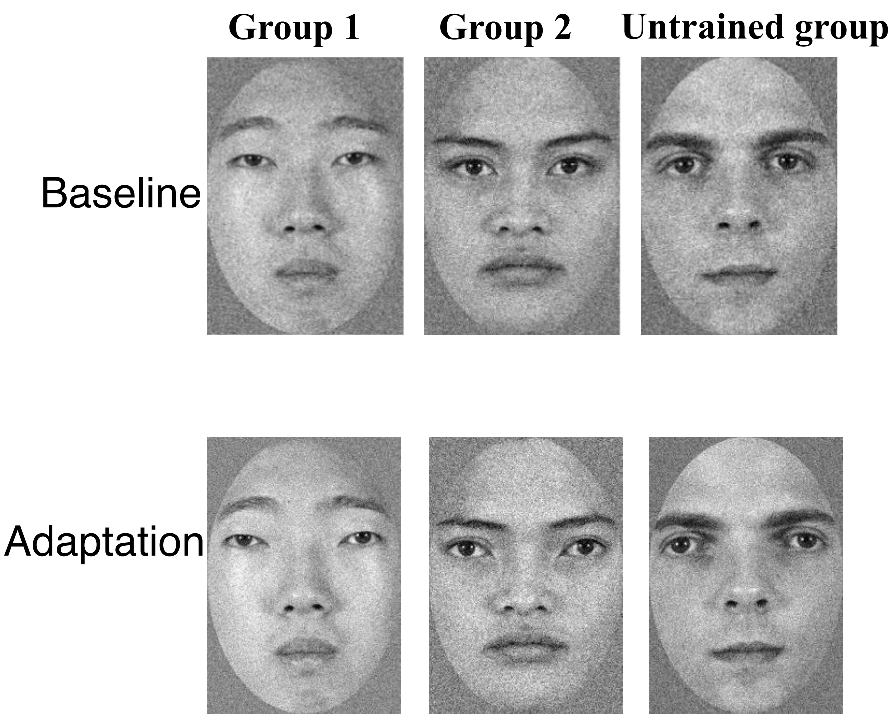

Figure 4.3 Adaptor faces used. The adaptor could either be a physically undistorted (baseline blocks) or a distorted face (adaptation) and varied across the 3 face groups.

Normality judgment task

Figure 4.4 shows test faces varied in eye separation, that were used in the normality judgment task. In the task, observers compared the normality of a test face with narrower-than-normal eye distance to another with wider-than-normal eye distance, and judged which appeared more normal (less distorted). The face with wider-than-normal eye distance was fixed ("fixed test face"), i.e. physically the same face throughout the experiment, while the face with narrower-than-normal eye distance varied ("varied test face") depending on observer responses. The fixed face is indicated with a black box in Figure 4.4.

We used the FaceAware Liquify tool in Adobe photoshop to create the set of test faces, with successive images varying in eye distance by 0.02 deg steps. We based this task on a similar one from prior work (Little et al., 2008a).

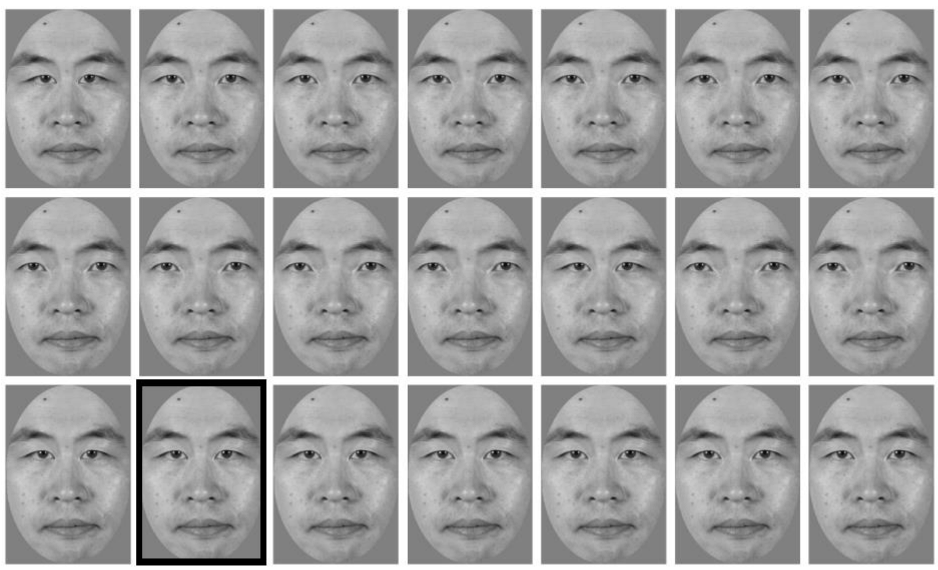

Figure 4.4 Test faces. Eye distance variation was 0.02 deg between successive test faces. The "fixed test face" is indicated by the black border.

Figure 4.5 shows an overview of one trial of the normality judgment task. In each trial of the normality judgment task, a narrow-eye test face ("varied test face") appeared on the screen for 500ms. The exact location of the test faces was slightly jittered 0.2 deg from the center of the screen, in both X and Y directions from trial to trial. This was done to mitigate any aftereffect (e.g. afterimage, adaptation aftereffect, etc.) that is retinotopic in nature. After this, a 4.5 secs "top-up” period followed, where observers viewed 3 presentations of the adaptor face (different adaptation conditions are described below). After top-up adaptation, the wide-eye distorted face (“fixed test face”) appeared for another 500ms. This was followed by 2 secs of blank screen during which observers responded by pressing a number key, ‘1’ or ‘2’, to indicate which of the two distorted test faces appeared to be more normal. A fixation marker was always shown at the center of the screen during the task, and changed its shape from square (adaptation) to circle (test) to help observers distinguish adaptor faces from test faces.

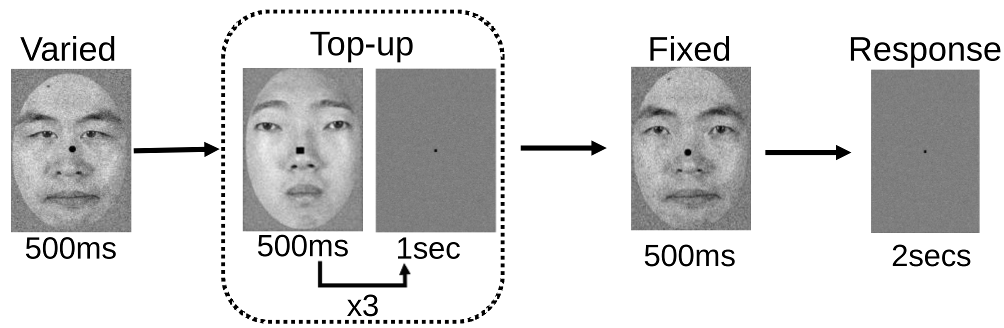

Figure 4.5 Overview of the normality judgment task. In each trial, the varied test face first showed up for 500 ms, followed by 3 ‘top-up’ presentations of the adaptor face (condition 1 adaptor shown here) for 500 ms, interleaved with 1 sec of blank screen. The fixed test face was then displayed for another 500ms, this time followed by 2sec blank screen duration during which observers responded whether the varied or the fixed test face appeared more normal. A square fixation marker was shown during adaptation, and a circular fixation marker was shown during test face presentation.

An adaptive ‘one-up-one-down’ staircase procedure determined which test face each observer saw from trial to trial. For example, if observers perceived the test face that had wider than normal eye distance (fixed test face) as more normal, compared to the test face that had narrower than normal eye distance (varied test face), then on the next trial the varied test face would have slightly wider eyes, and vice versa. Using this adaptive procedure, the varied test face eventually converged to a level of distortion that was perceptually equal to the distortion of the fixed test face. Such a face is defined as individual's point of subjective equality (PSE).

Adaptation Conditions

Observers were tested in different adaptation conditions using this paradigm. Trials were conducted in blocks of 30, where observers repeatedly viewed the same adaptor face. In 'baseline' blocks the adapter was an undistorted face. In adaptation blocks the face was distorted by shifting the eyes of an undistorted test face outward by 0.1 deg. In each session a baseline block was followed by an adaptation block.

Each block began with 1 minute of adaptation. During this period, the adapter face was presented for 500 msec every 1.5 sec, and we added a secondary task to control attention: Observers were instructed to press any number key if they detected the adaptor face appeared to be 'winking'. We edited the adaptor face in photoshop so that some presentations of the distorted face appeared to be winking. This secondary task was added to ensure that observers attended to the adaptor face during the 1 minute presentation, as attention can modulate adaptation aftereffects (e.g. in tilt (Spivey & Spirn, 2000), motion (Boynton, Ciaramitaro, & Arman, 2006), and contrast (Pestilli, Viera, & Carrasco, 2007) adaptation aftereffects).

Following the 1-minute adaptation period, the block of trials began, and in the middle of each trial, an additional 3 presentations of the adapter (500 msec each by 1sec blank screen) were shown. The presentation "topped-up" adaptation, kept it in a relatively constant state throughout the block.

Observers completed 30-trial baseline and 30-trial adaptation blocks in one session. Each observer was exposed to the three different adaptor faces in different sessions, with order counterbalanced across observers. Observers completed 12 sessions in total, two sessions for each of the three adapters, both before and after completing the visual search training.

Analysis procedure

To formally estimate observers' PSEs, we fit a Weibull function to their responses, which has the general form:

$$y = 1 - e^{-(\frac{x}{\alpha})^{\beta}}$$

where y is the fraction of the varied test face perceived as "more normal", x is the physical eye separation, and α and β are free parameters estimated by the fitting procedure. We used non-linear fitting that minimizes sum of squared residuals, and observer's PSE was estimated as the eye separation of the varied test face that appeared to be more normal than the fixed test face 50% of the time. Using this fitting procedure, we obtained each observer's PSEs for both the baseline and the adaptation blocks.

Rating task

Figure 4.6 shows a sample trial of the rating task. Observers' task involved rating, on a scale of 1-5, how similar any given pair of faces appeared to them.

Each observer viewed and rated 105 face pairs in total (15 faces, choosing two faces at a time). Details of this task were identical to those used Chapter 3.

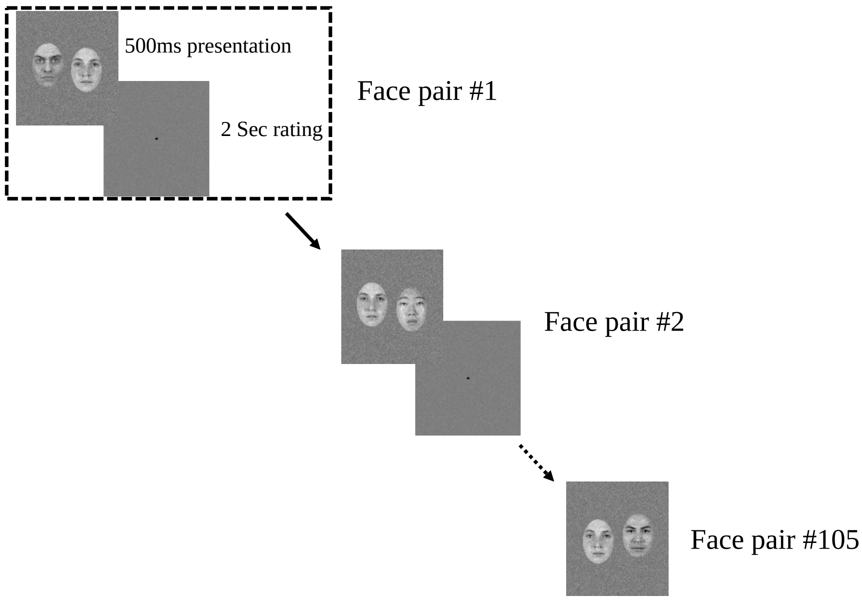

Figure 4.6 The rating task. On each trial, observers were shown randomly selected pairs of faces, one pair at a time for 500 msec, and observers had 2 seconds to respond, on a scale of 1-5, how likely they believed the face pair to be siblings.

Visual Search task

Figure 4.7 shows a sample trial of the search task. Observers were instructed to locate the target face, among 15 other distractor faces. Details of this task were identical to those used Chapter 3**.**

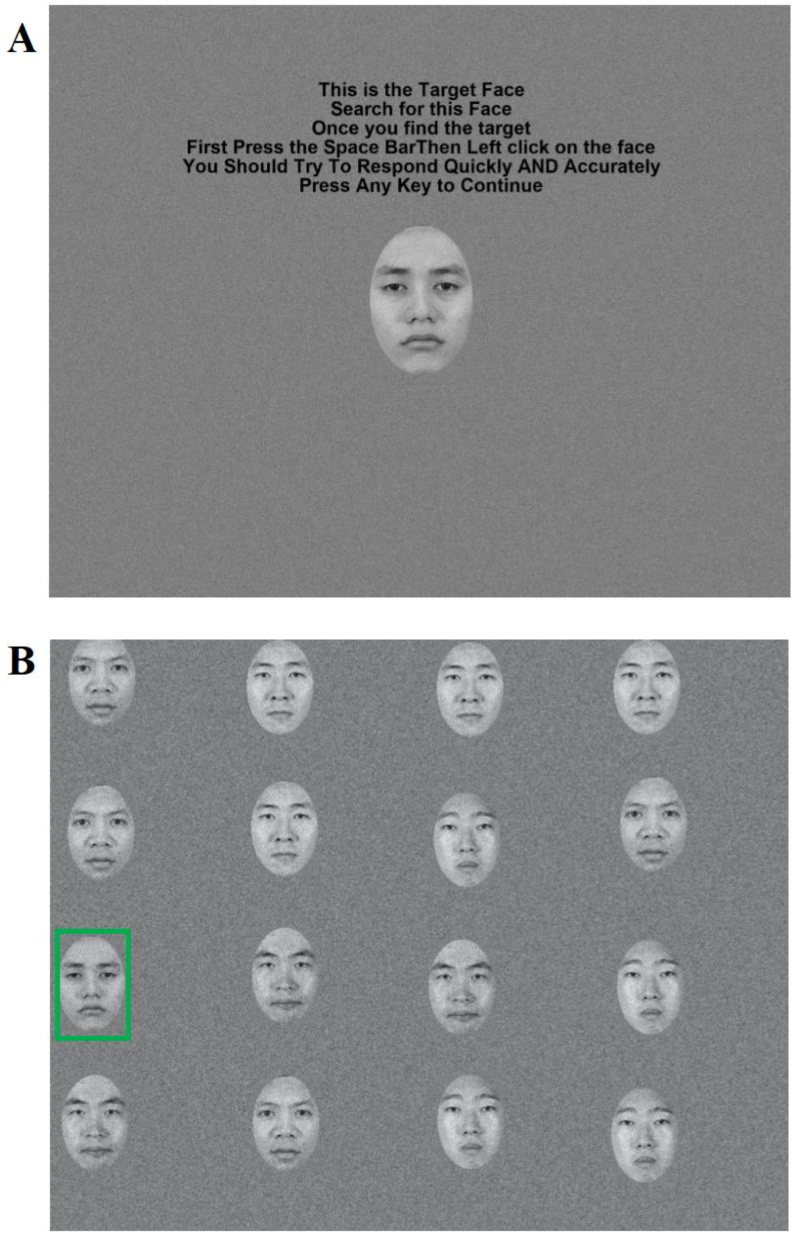

Figure 4.7 The search task. A: Observers were shown a target face and were instructed to locate this face as quickly and as accurately as possible. B: The 16 by 16 search grid.

Results

Visual search task

Over 3 days of training, observers became faster at locating target faces among distractors. Figure 4.8 shows the average search time across 3 days, for 10 observers; similar to the results we observed in Chapter 3, observers took significantly less time to correctly locate the target face during the final day of training (day 3) than during day 1. Average search time decreased by 0.6 secs across 3 days of training ($t(9)=-4.86$, $p<0.001$); this decrease in search time was slightly less than the decreasing trend we observed in Chapter 3 (1.02 seconds).

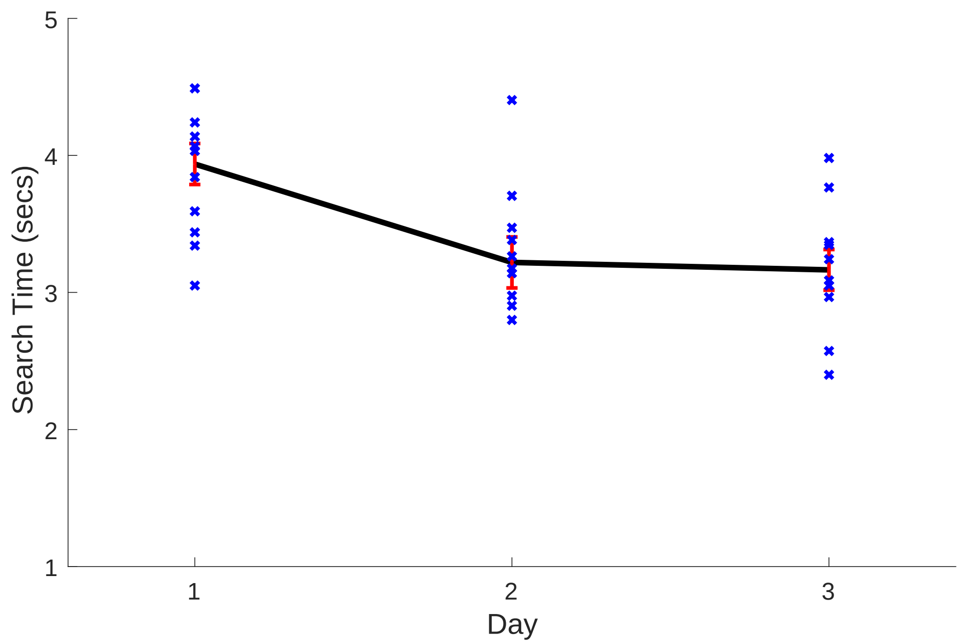

Figure 4.8 Average search time across three days of training (2 sessions per day). Black line plots the average (n=10) of the median reaction time across 3 days. Blue crosses show individual median reaction time across 3 days. Error bars represent S.E.M within a day

Similarity rating task

Before training

We analyzed results of the similarity rating task as in Chapter 3. Figure 4.9A shows the average dissimilarity matrix; figure 4.9B plots average perceptual dissimilarity ratings for "within group" and "between group" face pairs. Again, a larger dissimilarity score indicates face pairs became **less similar**.

As in Chapter 3, observers were sensitive to the physical similarity criteria used to define the three face groups. Overall, faces that belonged to the same group appeared more similar to each other than to faces of another group. Specifically, observers' dissimilarity ratings for within group face pairs were lower than for between group face pairs (mean difference in grouping index = 1.1 ($t(9)=9.16$, $p<0.001$) (comparing left and right sets of bars in Figure 4.9).

Additionally, observers were again sensitive to the physical similarity between the two trained groups, compared to the untrained group. Specifically, dissimilarity ratings for "between-group" face pairs differed significantly between the three compared groups ($F(2,18)=21.41$, $p<0.001$). Pairwise comparisons showed that dissimilarity between face pairs that belonged to different trained groups (1&2) was significantly less than dissimilarity between pairs where one face belonged to the trained groups and the other from the untrained group (difference in dissimilarity = -0.85 ($t(9)=-4.137$, $p<0.005$) (compare first bar on right of Fig 4.9B to adjacent two).

Also as in Chapter 3, observers' ratings of dissimilarity for face pairs within the same groups did not differ significantly across groups ($F(2,18)=0.884$, $p>0.05$) (comparing left sets of bars in Figure 4.9).

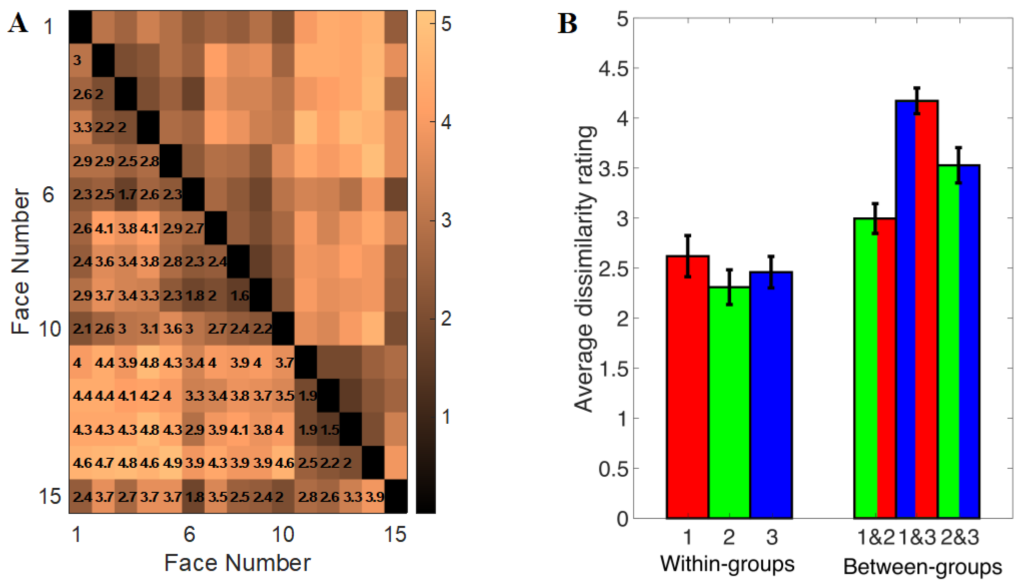

Figure 4.9 Dissimilarity ratings for all possible face pairs, before training. A: Average dissimilarity rating matrix for all possible face pairs. Higher rating scores correspond to face pairs appearing more dissimilar. B: Average rating scores for within and between-group face pairs. X-axis labels, '1','2' and '3', correspond to face pairs within trained groups 1, 2 and untrained group 3.

<u>After training</u>

Training again helped observers distinguish faces within the trained groups. Figure 4.10A shows the dissimilarity matrix, and Figure 4.10B shows average ratings for within group and between group face pairs. After 3 days of training, within-groups dissimilarity ratings differed significantly across groups ($F(2,18)=3.26$, $p<0.05$). Faces within the trained groups on average appeared more dissimilar than faces within the untrained group (difference between trained vs untrained = 0.52, $t(9)=2.63$, $p<0.05$).

Similar to what we saw in Chapter 3, even after 3 days of training observers still appeared to be sensitive to physical differences that we used to define groups: Observers' dissimilarity ratings for within group pairs were lower than for between group pairs (grouping index = 0.98, (t(10)=7.81, p<0.001).

Nevertheless, as in Chapter 3, two trained groups remained more similar to each other than to the untrained group. Dissimilarity ratings for between group face pairs was lower for the two trained groups, compared to when one of the faces was untrained (comparing right sets of bars in Figure 4.10B). Specifically, between group face pairs differed significantly across the 3 groups ($F(2,18)=16.11$, $p<0.001$). Pairwise comparisons showed that dissimilarity between face pairs that were from trained groups 1&2 was reliably less than faces pairs where one was in a trained group and the fellow was in the untrained groups (mean difference in dissimilarity = 0.6, $t(9)=3.53$, $p<0.005$).

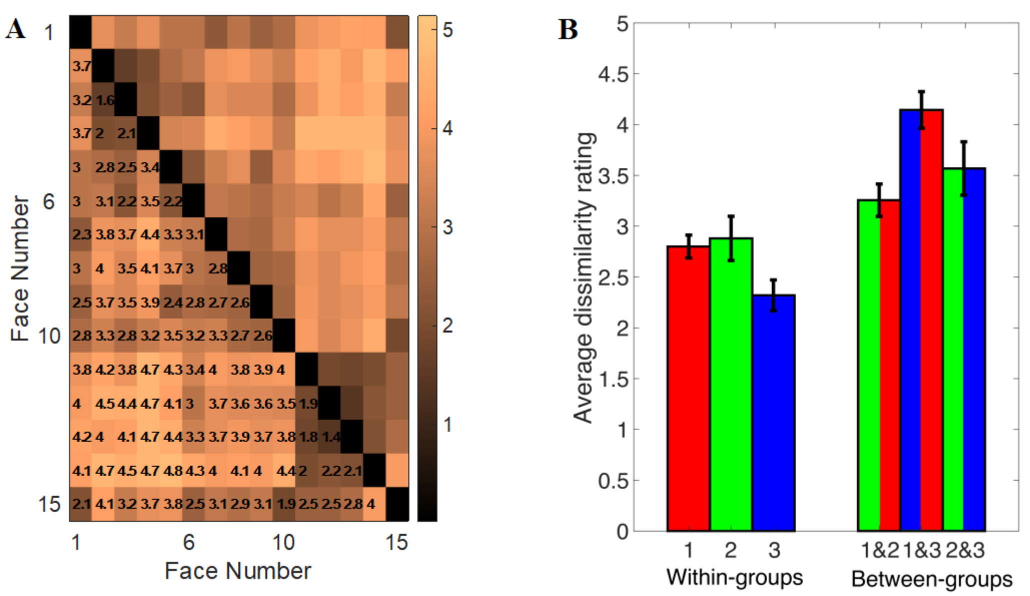

Figure 4.10 dissimilarity rating for all possible face pairs, after training. 10B: After training average rating scores for within and between-group face pairs.

Comparing dissimilarity ratings before and after training

Overall effects of training were smaller than in Chapter 3. Figure 4.11 shows average dissimilarity ratings both before and after training. Training increased dissimilarity for face pairs within a group ($F(2,18)=1.26$, $p<0.01$); however, this effect was only significant for trained group 2 (difference in dissimilarity = 0.57, $t(9)=2.74$, $p<0.05$). Dissimilarity of face pairs within trained group 1 or the untrained group did not change reliably ($p>0.05$).

Dissimilarity between the two trained groups increased following training: Face pairs with one face in each trained group increased in dissimilarity by 0.26, although this increase was not statistically significant, $p>0.05$. Dissimilarity for pairs where one face was from trained group and one from an untrained group, increased by only 0.006 ($p>0.05$). This increase in dissimilarity was lower than the increase in dissimilarity of face pairs between the two trained groups, mean difference=0.25, but this difference was also not statistically significant ($p>0.05$).

Again in this study, after 3 days of training observers were relatively equally sensitive to the physical differences of faces across the predefined groups: the grouping index did not change much after training compared to before training (change in grouping index = 0.11, $p>0.05$).

Comparing before and after training similarity ratings shows that 1) face pairs within the same trained groups appeared less similar following training, although this change was not significant for members in trained group 1; 2) face pairs between the two trained groups ('1&2' in figure 4.11) also appeared less similar following training, although this change was again not statistically significant; 3) similarity within the untrained group and between trained and untrained groups ('1&3' and '2&3' in Figure 4.11) changed relatively little. These results follow the same general pattern as in Chapter 3, suggesting the establishment of group representations for the trained groups, however, most effects did not reach statistical significance.

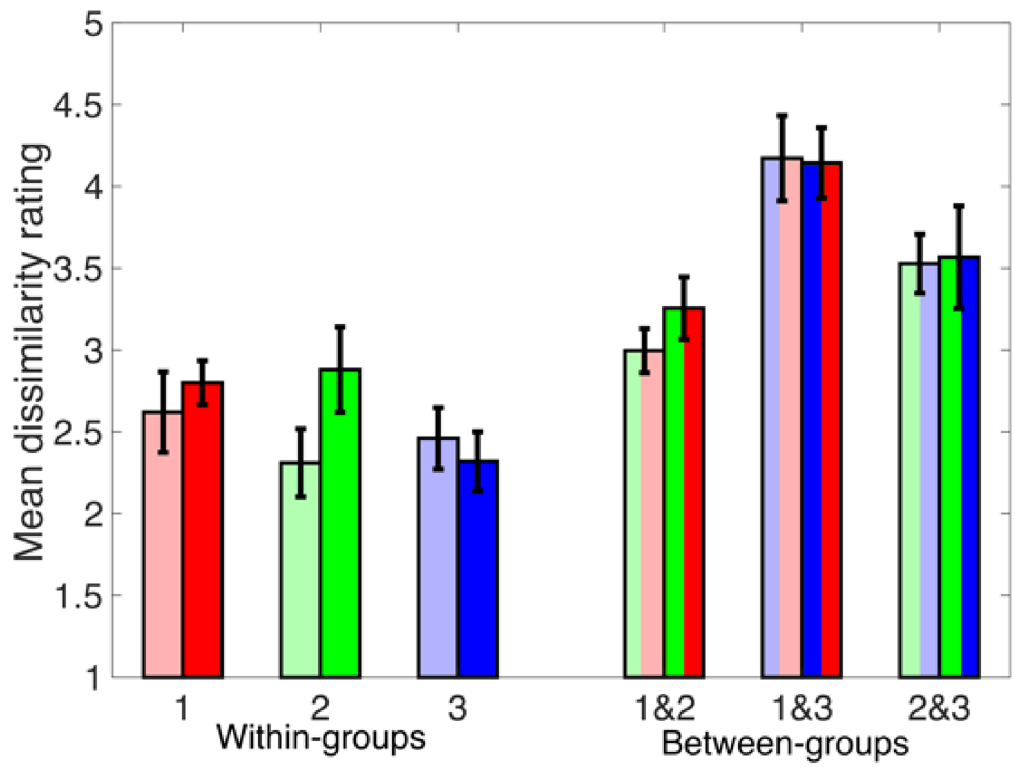

Figure 4.11 Average dissimilarity ratings for within and between-group face pairs, before and after training. Bar colors represent the 3 pre-defined groups; saturation of the colors represents average similarity rating before (less saturated) and after training.

Hierarchical clustering analyses of dissimilarity ratings

Figure 4.12 visualizes hierarchical clustering results for dissimilarity scores before and after training. To identify clusters of faces that appeared similar prior to and after 3 days of training, we again conducted an agglomerative hierarchical clustering analysis: Pairs of faces that appeared most similar were first grouped into clusters, and this process repeated until all faces were merged into one single cluster.

Prior to training, the dendrogram reveals evidence for both two and three clusters of faces, similar to what was seen in Chapter 3. Two clusters form at the top of the dendrogram and remain stable as one moves down the dendrogram, until the trained faces split into two groups. Moving further down the tree, these three clusters rapidly divided into many smaller clusters. However, there is also evidence that some faces (specifically faces 1, 6 and 15) were not strongly clustered in their pre-defined group. Potential reasons for this will be discussed below.

Following training, again starting from the top, there is evidence for two clusters comprising the trained and untrained faces; however, this grouping, as in Chapter 3, relatively quickly breaks into three groups. Moving further down the dendrogram, these 3 clusters again further divided into smaller clusters.

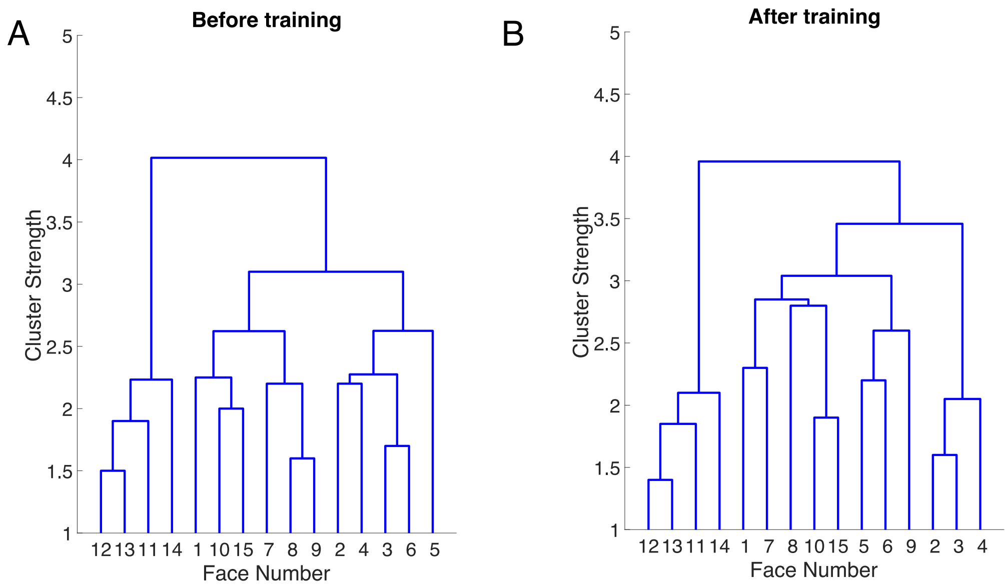

Figure 4.12 Dendrograms from clustering analyses of before and after training dissimilarity ratings. Faces were arranged based on pairwise similarity on the x-axis; y-axis plots clustering strengths, which measures the average perceptual dissimilarity of faces within clusters; smaller numbers indicate stronger clustering.

To formally evaluate the evidence for different clustering solutions, we again computed our standard grouping index, the difference of between- and within-cluster pairwise dissimilarities, for two- and three-cluster solutions. Figure 4.13 plots the grouping indices for two- and three-cluster solutions before and after training. A larger grouping index indicates stronger clustering of faces.

Prior to training there was relatively equal evidence for 2- and 3- group representations, and after training the 3-group solution dominated. Before training, difference in grouping indices between the 3-cluster and the 2-cluster solution was not reliably different (difference in grouping index= 0.15, $p>0.05$). After training, the three-cluster solution yielded a reliably larger grouping index (difference= 0.17, $t(9)=2.94$, $p<0.01$). However, the interaction between training and clustering solution was not significant ($p>0.05$).

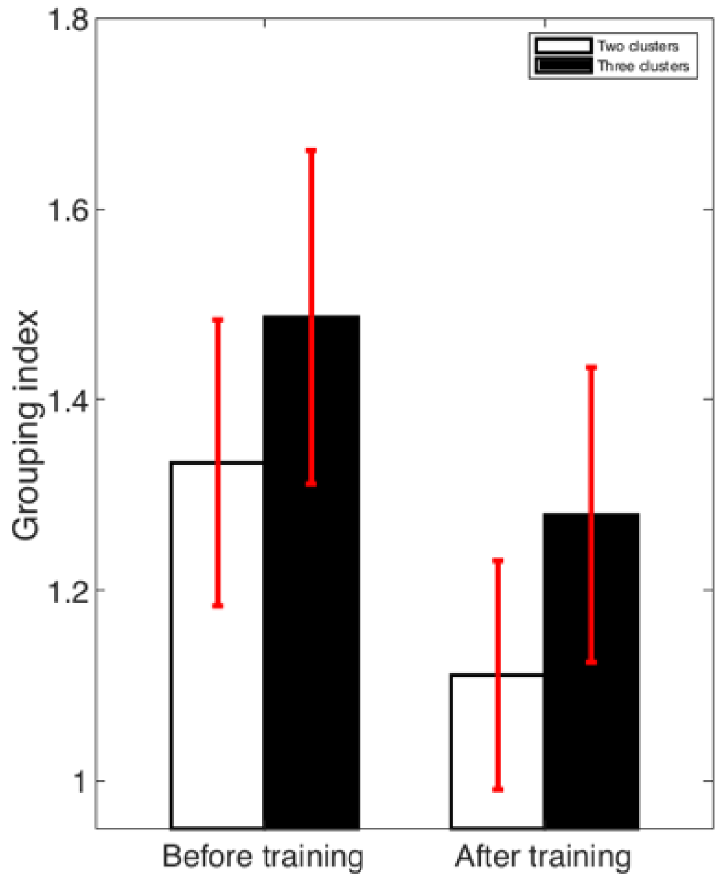

Figure 4.13 Grouping indices for two- and three-cluster solutions, before and after training. Error bars represent S.E.Ms of the grouping indices for n=10 observers, for different clustering solutions before and after training.

Figure 4.14 shows the perceptual dissimilarity matrices rearranged to reflect the clustering solutions. In figure 4.14A (before training dissimilarity matrices), pre-defined trained groups of faces (faces 1-10) and untrained face 15 were tightly clustered, and the other untrained faces (11-14) formed a distinct cluster. Following 3 days of training, there were at least 2 distinct clusters now formed among trained faces: faces 1,7,8,10, 5,6,9 and untrained face 15 were perceptually more similar than other faces, as well as faces 2,3 and 4. Untrained faces 11-14 formed a single, tight cluster before and after training.

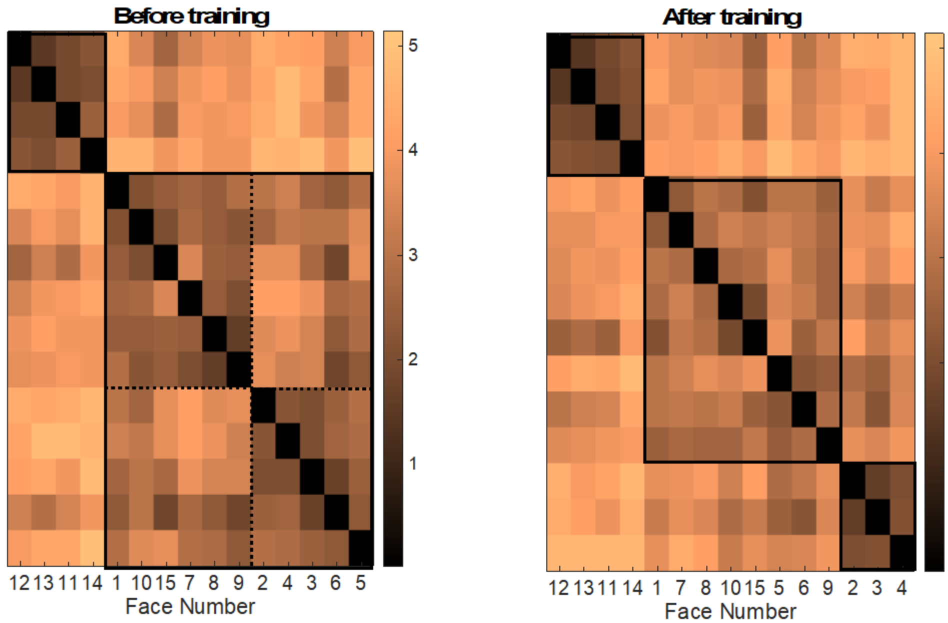

Figure 4.14 Dissimilarity matrices, with clusters of faces enclosed with rectangles: Before training, solid lines show a 2-cluster solution, and dashed lines show the alternative 3-cluster solution. X-axes of the matrices were rearranged so that faces belonged to the same were close to each other on the X-axes.

Face adaptation task

<u>Before training</u>

To measure face adaptation, we used the PSE in the normality judgment task- i.e. the closed-eye face that appeared equally distorted to the fixed wide-eyed face (see Methods). The face adaptation aftereffect (FAE) was measured by taking the difference in PSEs between baseline and adaptation. Because the adapter was always a wide-eyed face, standard results predict an increase in PSE, as adaptation is expected to make wide-eyed faces appear more normal. Observers adapted to wide-eye distorted faces that came from the three pre-defined groups in different sessions, and the face adaptation effect in all three conditions was measured with a test face that belonged to trained group 1.

Here, we report PSE as percentages relative to the physical distortion in the fixed test face (Fig 4.15), which was 0.55 deg. 0% change corresponds to the varied (close-eyed) test face that is physically as distorted as the fixed test face (i.e. eyes closer by 0.55 deg than the undistorted face. All other values are reported as percent change scores relative to this distance of 0.55 deg. Thus, the eyes in the fixed test face were 35% wider than the eyes in the face at the zero point.

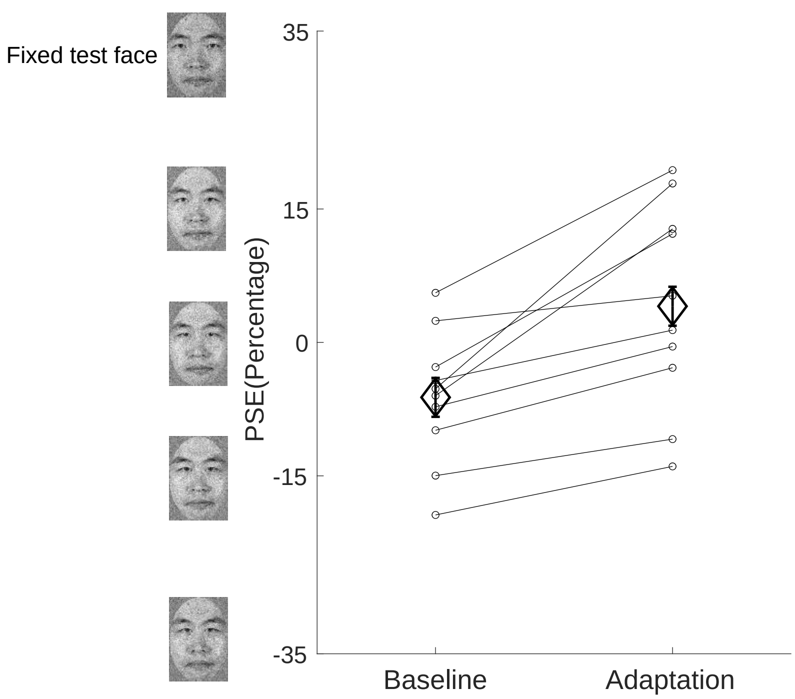

Figure 4.15 An illustration of PSE in percentages and its relationship with eye separation of the varied test face. 0% corresponds to eye separation of the varied test face that is as distorted physically distorted as the fixed test face. Positive percentage changes correspond to varied test faces with wider eye separation, and negative percentages with narrower eye separation.

Figure 4.15 showed observers' PSE before and after adapting to the group 1 adaptor. During the baseline, observers on average perceived the fixed test face more distorted as it was: Average PSE in baseline blocks across the three adaptor faces was 6.5% more distorted than the test face physically as distorted as the fixed test face (i.e. a value of -6.5% in the Baseline condition).

In general, adaptation to a wide-eyed face increased PSEs in the expected direction, making wide-eyed faces appear more normal (Fig 4.16): Average PSE increased by about 7.14% after adaptation across all three adaptors ($F(2,18)=14.31$, $p<0.001$).

As expected, however, adaptation was selective, with adapted PSEs differing significantly across adaptors. Specifically, the FAE was strongest when the adaptor was from trained group 1, the condition where the adaptor and the test faces both belonged to the same group (PSE increased by 10.24%, $t(9) = 6.09$, $p<0.001$). Additionally, the FAE was also significant when the adaptor was from group 2 (PSE increased by 6.21%, $t(9) = 3.36$, $p<0.01$). This is evidence that effects of adaptation transferred from group 2 (where the adapter was located) to group 1 (where the test was). Lastly, the FAE was not significant when the adapter was from group 3 ($p>0.05$), indicating lack of reliable transfer of adaptation in this case. Multiple pairwise comparisons revealed that the difference in FAE was the largest between adaptors from groups 1&3 (difference in PSE=5.25%, $p<0.001$). The difference in PSE between adaptors from groups 1&2 was intermediate in size (4.03%, $p<0.001$); Lastly, the difference in PSE between adaptors from groups 2&3 was not significant ($p>0.05$).

These results show that before 3 days of training: 1) The FAE was the strongest when the adaptor was from group 1, the same group as the test face; 2) The FAE transferred between faces that belonged to different trained groups; this is evidence that these two physically similar groups share representation to some extent. 3) The FAE did not reliably transfer between trained group 1 and the untrained group.

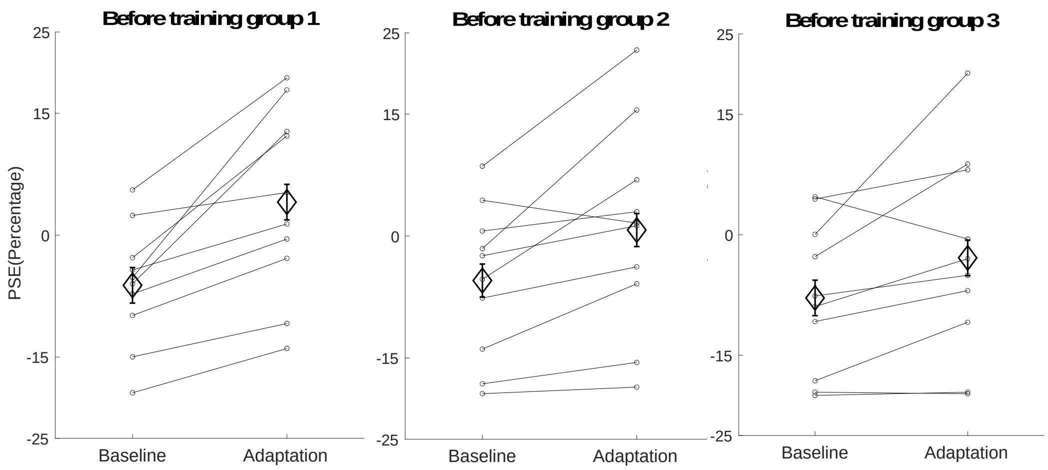

Figure 4.16. PSEs across the 3 pre-defined groups of adaptor faces before training. Individual observers' PSEs, inter-eye distance of the varied test face that appeared equally as distorted as the fixed test face, are plotted with the group means, measured both during baseline and post-adaptation. Error bars represent the standard error of the mean difference score (as appropriate for a within-subjects design, Loftus & Masson, 1994). The dotted line represents at 0% represents an inter-eye distance of the face that is physically as distorted as the fixed test face. Positive change corresponds to a shift of PSE to a face that has wider eyes.

<u>After training</u>

Again, adaptation generally increased PSEs in the expected direction (Figure 4.17): Average PSE increased by 5.81% after adaptation across all adaptors ($F(2,18) = 9.29$, $p<0.005$).

Adaptation also showed specificity, differing in strength across adaptors: Multiple pairwise comparisons showed that the FAE was still the strongest when the adaptor was from trained group 1 (PSE increased by 8.63%, $t(9) = 4.82$, $p<0.001$); however, the FAE was no longer significant when the adaptor was from group 2 (PSE increased by 4.80%, however $p>0.05$). Finally, similar to before training, the FAE was not significant for group 3 adaptors ($p>0.05$). Multiple pairwise comparisons revealed that the FAE was the largest between adaptors from groups 1&3 (difference in PSE=4.63%, $p<0.005$); difference in PSE between adaptors from groups 1&2 was 3.82%, $p<0.01$; Lastly, difference in PSE between adaptors from groups 2&3 was not significant ($p>0.05$).

When comparing FAE before and after training across the three adaptors, adaptation decreased in all three conditions: The adaptor from group 1 showed the largest change in FAE after learning, decreasing by 1.61%, followed by adaptor from group 2 (1.40%) and adaptor from group 3 (0.97%). However, these changes were not significantly different from each other ($F(2,18) = 0.034$, $p>0.05$).

In sum, after 3 days of training: 1) FAE was still the strongest when the adaptor and the test belonged to same trained group, suggesting shared representation between the adaptor and test; 2) however, after training FAE did not reliably transfer between faces that belonged to different trained groups. This indicates at least partial evidence for dissociable group-level representations for the adaptor and test faces, which belonged to different trained groups; 3) The FAE did not transfer between the trained and the untrained group.

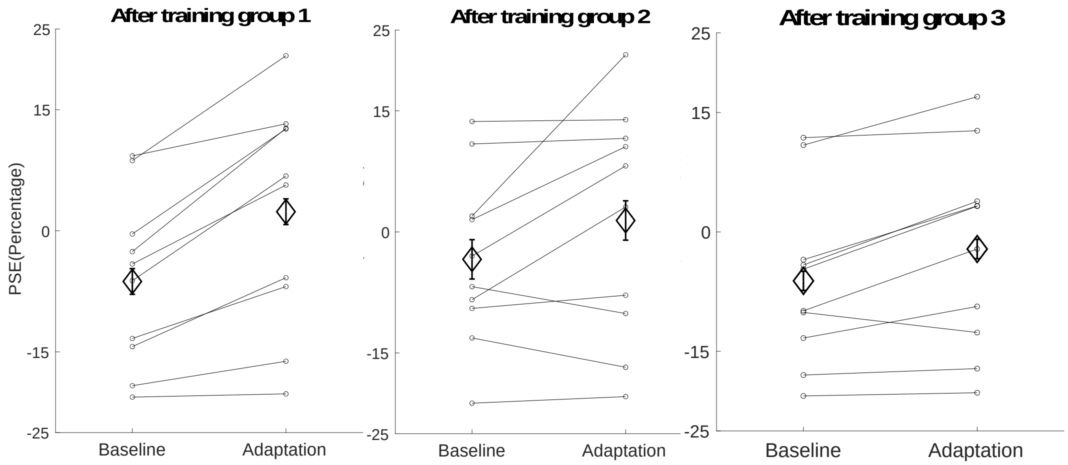

Figure 4.17. After training PSEs across 3 adaptor groups. Plotting conventions are as in Figure 4.16.

Comparing results across Asian and non-Asian observers

Face representation can be influenced by both short-term and long-term exposure to the surrounding face environment. Results from Chapters 3 and 4 showed that short-term (3 days) training helped observers established group-level representations for face pairs of different pre-defined groups. To examine whether long-term exposure to certain face environments interacts with our short-term manipulation, we compared results gathered in Chapters 3 and 4 across Asian and non-Asian observers, under the assumption that these groups differed in relative experience for encoding the Asian faces that comprised our trained groups.

16 observers participated in Chapter 3 studies, and 10 in Chapter 4, and the number of Asian and non-Asian observers were evenly split in all experiments. We combined the dissimilarity rating data reported in both chapters and compared Asians and non-Asians' responses (n=13 both groups).

Dissimilarity rating

Before training, Asian observers found face pairs within the same trained groups, as well as pairs between different trained groups to appear slightly more dissimilar than non-Asian observers (Fig 4.18A). The mean dissimilarity rating for pairs within the same trained group was 2.36 for Asian observers, and 2.17 for non-Asian observers (larger number means more dissimilar); mean dissimilarity rating for pairs between trained group was 2.88 for Asian observers, and 2.67 for non-Asian observers. However, these differences were not statistically significant, $p>0.05$.

Throughout chapters 3 and 4, we defined the grouping index by comparing dissimilarity ratings of face pairs across different trained groups versus pairs within the same trained group, and used this index as a measurement for grouping strength. Before training, the average grouping index was 1.23 (Fig 4.18A differences between left and right sets of bars) for Asian observers and 1.15 for non-Asians, however this difference was also not significant.

After training, grouping index was 1.27 for Asians and 1.37 for non-Asians; this difference was again not significant. However, 3 days of training increased the grouping index for non-Asian observers by 0.22, and only by 0.04 for Asian observers, and this difference was statistically significant ($t(12)=2.30$, $p<0.05$). This suggests training strengthened group-level representations of trained faces reliably more for non-Asian than for Asian observers.

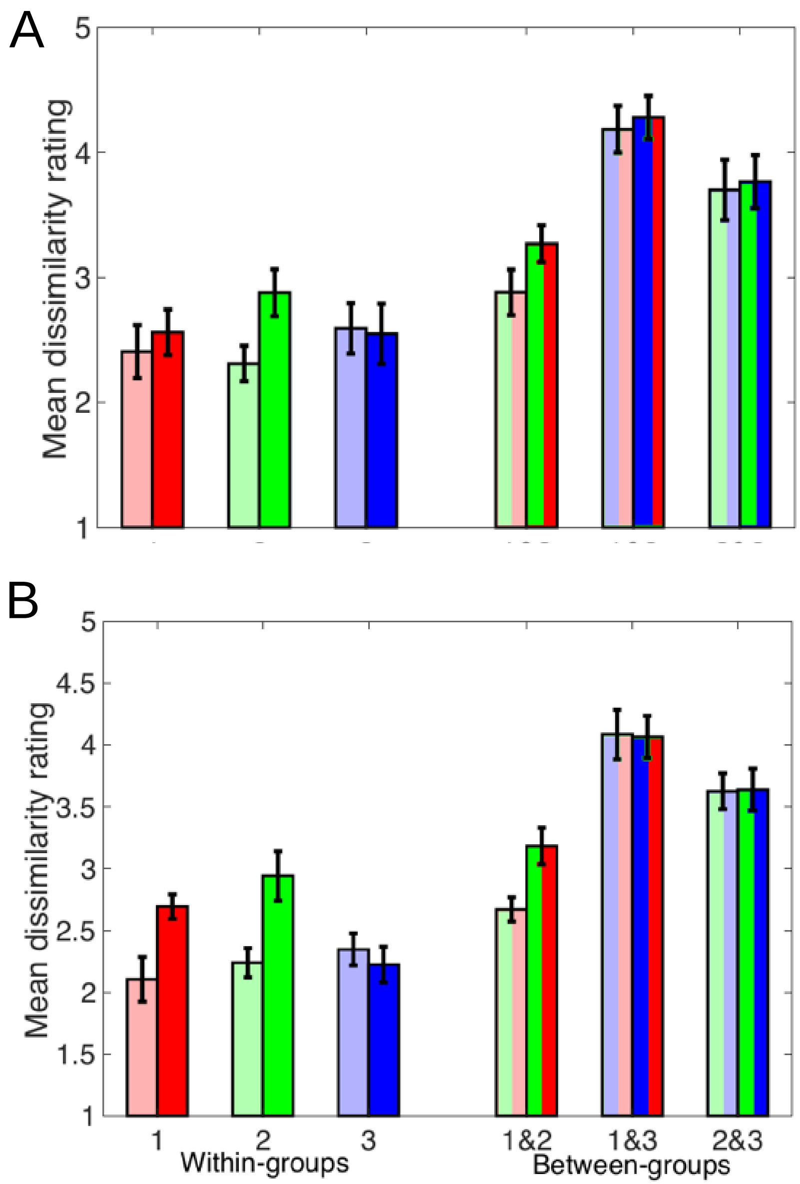

Figure 4.18 A: Average dissimilarity ratings for Asian observers. B: Dissimilarity ratings for non-Asian observers. Larger numbers correspond to pairs appearing more dissimilar

Clustering analyses

Hierarchical clustering results also differed across Asian and non-Asian observers. Dendrograms of Asians (Fig 4.19A) and non-Asians (Fig 4.19B) before and after training show that while there was equal evidence for either two or clusters of faces for Asian observers before and after training, dendrogram of non-Asian observers after training showed a much clearer trend of trained faces further dividing up into smaller clusters faces 1-5, faces 7,8,10 and faces 6,9.

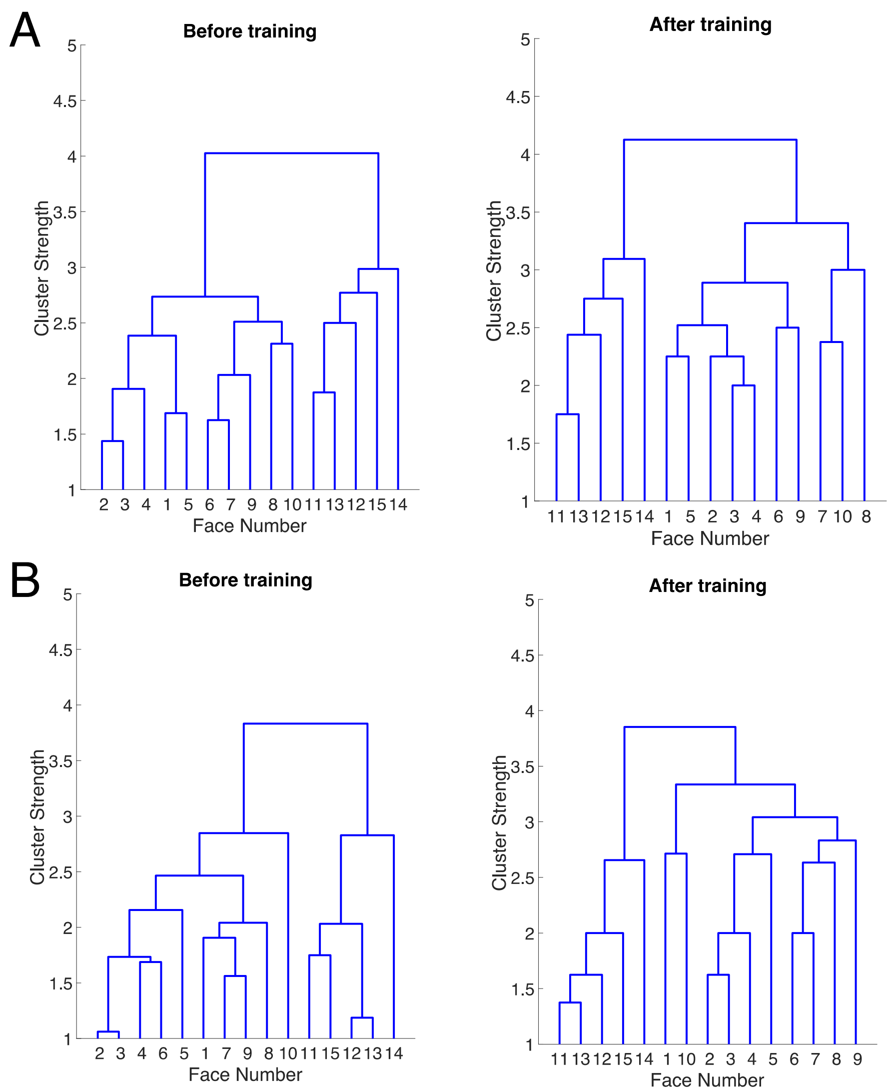

Figure 4.19 A: Before and after training dendrograms for Asian observers. B: Dendrograms for non-Asian observers

To further examine evidence for different clustering solutions, we examined the grouping index (Fig 4.20). Before training, the index did not differ reliably for either non-Asians (difference in grouping index between 2- and 3-cluster solutions = 0.05, $p>0.05$) or Asians (difference = 0.12, $p>0.05$). After training, overall evidence for a 3-cluster solution was stronger than for a 2-cluster solution, as grouping indices were larger for a 3- than for a 2-cluster solutions, for both Asian and non-Asian observers (Fig 4.20). Difference in grouping index between 2- and 3-cluster solutions was 0.18 for non-Asian observers ($p<0.05$), and 0.03 ($p>0.05$) for Asian observers, and this difference is statistically significant ($t(12) = 2.07$, $p<0.05$). Therefore, while prior to training dissimilarity ratings from both observer pools showed equal evidence for a 2- or a 3-cluster solution, there is greater evidence for a 3-cluster solution in non-Asian observers after training.

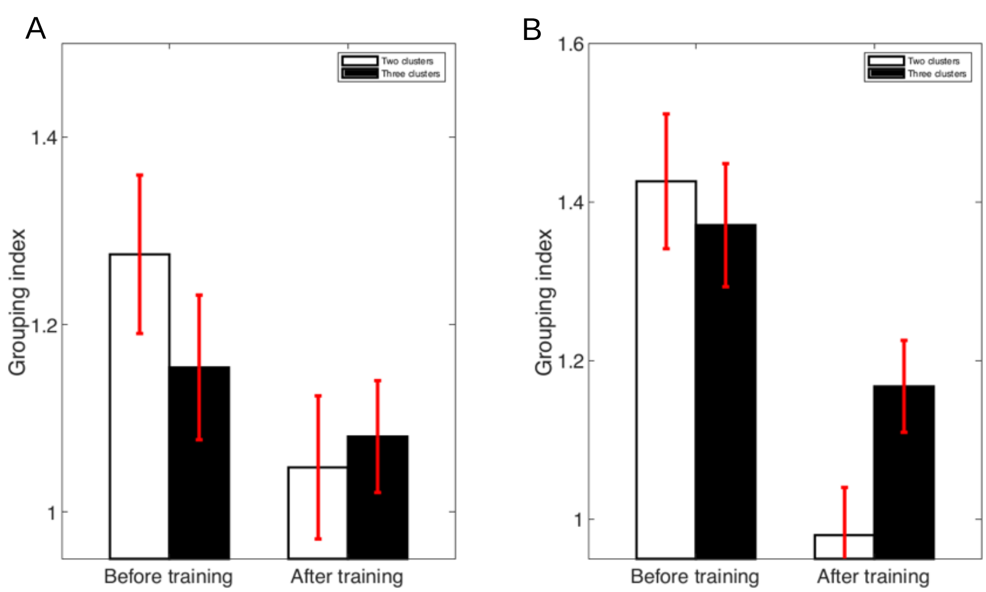

Figure 4.20 A: Asian observers' before and after training grouping indices of 2- and 3-cluster solutions. B: Non-Asian observers' grouping indices for different clustering solutions, before and after training.

<u>Transfer of the FAE</u>

Prior to training, adaptation to a wide-eyed face generally increased PSEs in the expected direction, for both Asian (average increase in PSE = 4.15%; $F(2,8)=3.67, p>0.05$; Fig 4.21A) and non-Asian observers (increase in PSE = 10.14%; $F(2,8)=7.07, p=0.05$; Fig 4.21B). Non-Asian observers showed overall a reliably stronger FAE than Asian observers (difference in PSE = 5.99%; $t(4)=2.56$, $p<0.05$).

Both Asian and non-Asian observers showed adaptation effects when the adaptor and test were from the same trained group (group 1 adaptor)as PSE increased by 7.89% for Asians ($t(4)=2.79$, $p<0.05$), and 12.59% for non-Asians ($t(4)=3.83, p<0.01$). However, adaptation showed greater specificity among Asian observers: Adaptation effects transferred between faces that belonged to different trained groups (group 2 adaptor) only for non-Asian observers (increase in PSE = 9.61%, $t(4)=3.07$, $p<0.05$), but not for Asian observers (increase in PSE = 2.81%, $p>0.05$). Lastly, neither Asian nor non-Asian observers showed adaptation effect when the adaptor was from the untrained group (group 3 adaptor).

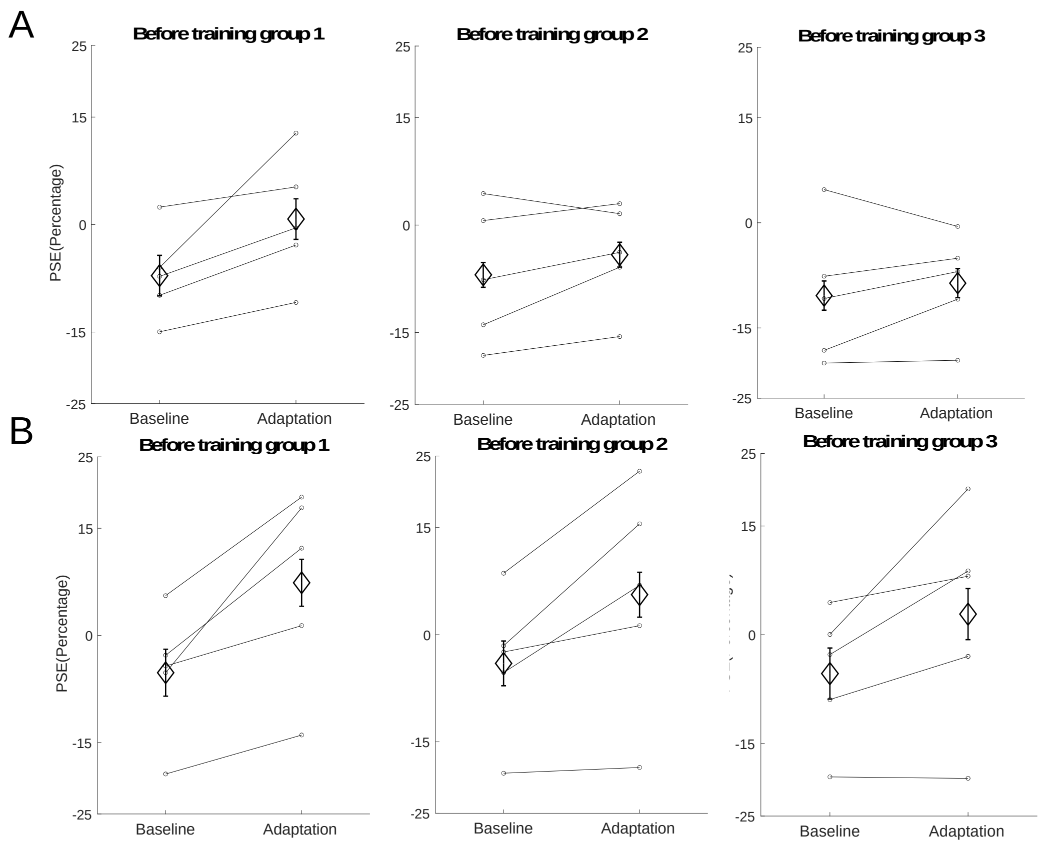

Figure. 4.21 Before training PSEs across the 3 pre-defined groups of adaptor faces. A: Asian observers. B: Non-Asian observers

After training, the overall FAE was again greater for non-Asian observers (average increase in PSE = 7.71%; Fig 4.22B) than for Asian observers (Fig 4.22A; PSE increased by 3.92%); this difference of 3.79% is again statistically significant ($t(5)=1.79$, $p<0.05$).

Both observer pools again showed adaptation effects when the adaptor and test were from the same trained group, as PSE increased by 6.45% for Asian observers ($t(4)=3.05$, $p<0.05$) and 10.81% for non-Asians ($t(4)=5.06$, $p<0.01$). However, after training adaptation now show specificity for both observer pools. When the adaptor and test were from different trained groups, PSE increased by 2.16% for Asian observers ($p>0.05$), and 7.45% for non-Asians ($p>0.05$). This suggests that training helped developed distinct representations for the two trained groups of faces. Lastly, neither observer pool showed adaptation effects when the adaptor was from the untrained group.

When comparing effects of training on the FAE across Asians and non-Asian observers, group 2 adaptor showed the biggest difference in PSEs between the two observer pools (1.52%), followed by the untrained group (1.45%) and group 1 adaptor (0.34%). However, none of these differences is statistically significant.

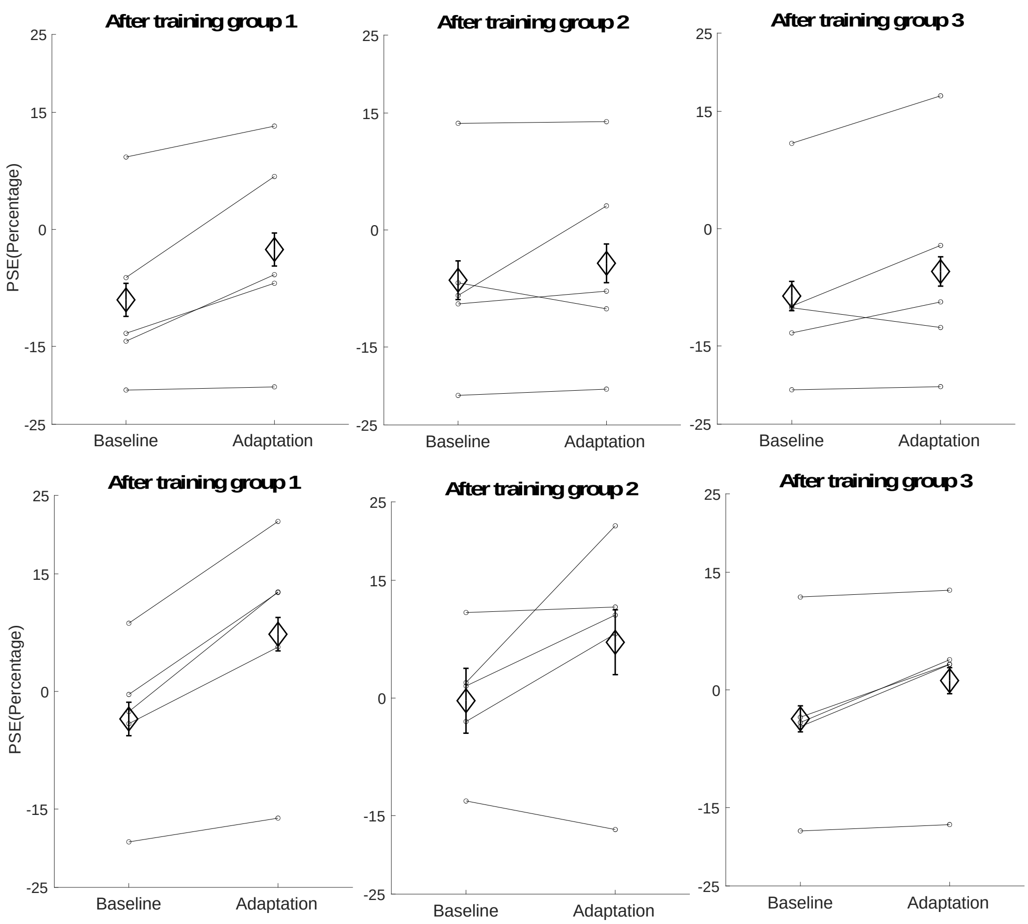

Figure 4.22. After training PSEs across the three adaptor groups. A: Asian observers. B: Non-Asian observers

Discussion

In this study we found that after 3 days of training: 1) Observers were significantly faster at locating target faces from distractors that appeared similar. 2) Similarity within and between the trained groups of faces decreased; clustering analysis showed that similar to what we observed in chapter 3, after training new clusters of faces were formed among trained groups. 3) Face adaptation aftereffects transferred less between the two trained groups, consistent with distinct group-level representations forming across the trained groups.

The decrease in transfer of adaptation aftereffects suggests that observers learned to establish more distinct representations for the groups of faces that we defined. Previous studies have showed that face adaptation aftereffects are contingent upon race (Jaquet et al., 2008), gender (Little, DeBruine, & Jones, 2005), age (Little, DeBruine, Jones, & Waitt, 2008), and facial expression (Campbell & Burke, 2009), identity (Soto, Escobar, & Salan, 2020). Our results build upon these previous findings to show that training can alter these contingencies and help separate representations of faces that initially might be represented under more shared coding mechanisms. An apparent advantage of forming these separate representations is that faces that initially appeared similar looked more distinct, which could improve observers' performance on perceptual tasks that involve face perception. The visual system could switch between these distinct representations depending on the surrounding 'face environment' to achieve a more accurate face perception.

One possible account of this improvement is that the visual system learned to utilize the ‘most useful features’ for differentiating among faces that initially appeared similar (Levin, 2000; Valentine, 1991). Future studies could use various methods to measure the apparent or latent facial features (could be a weighted combination of various facial features) that help form a more accurate representation of trained faces.

Increasing the amount of learning perhaps could better reveal how perceptual learning alters face encoding and representation. Although in this study the average reaction time data showed a decreasing trend across three days of learning, other literature suggested that observers continued to improve on the visual search task even after practicing the same task for over a period of one month (Collins, Behrmann, Tanaka, & Wong, 2020). For some of our observers that showed strong transfer of FAE, before and after learning, perhaps more learning is needed in order to establish separate representations between the two trained groups of faces that appear similar.

This study was designed to examine the dynamics of face representation; however, we have only investigated representation using behavioral methods. Imaging studies, on the other hand, have provided more direct evidence for dynamic face representation in the visual system: For example, neural responses of familiar faces were decodable in cortical regions that represent these faces (Afzalian & Rajimehr, 2021); in another study, neural responses to different images of the same familiar face were found to be largely invariant, in brain areas that encode familiar faces (Weibert et al., 2016). Future studies could measure how our group-based learning could modulate neural activity in brain region of interest, applying various univariate and multivariate approaches.

Our visual system is capable of rapidly adjusting to different perceptual environments by learning to maintain and switch between different “modes” appropriate for distinct color environments (Li et al., 2020). In human face perception, given faces are most likely encoded relative to a “norm” (see chapter 1 for a review on norm-based coding of faces). Our results suggest that there may be partially or fully independent ‘sub-spaces’ for faces from distinct groups; each subspace could maintain its own norm face, and our visual system might implement various computational and neural processes to determine the most optimal sub-representational space to use, given the perceptual environment. For example, our visual system could utilize ensemble coding by computing a summary statistical representation of all faces encountered in an environment (e.g. Haberman & Whitney, 2009), and switch to the “most appropriate” representational space. Importantly, such a dynamic representation might generalize beyond face and color perception.

Bibliography

Afzalian, N., & Rajimehr, R. (2021). Spatially Adjacent Regions in Posterior Cingulate Cortex Represent Familiar Faces at Different Levels of Complexity. https://doi.org/10.1523/JNEUROSCI.1580-20.2021

Apps, M. A. J., & Tsakiris, M. (2013). Predictive codes of familiarity and context during the perceptual learning of facial identities. *Nature Communications 2013 4:1*, *4*(1), 1–10. https://doi.org/10.1038/ncomms3698

Barrett, S. E., & O'Toole, A. J. (2009). Face adaptation to gender: Does adaptation transfer across age categories? *Visual Cognition*, *17*(5), 700–715. https://doi.org/10.1080/13506280802332197

Boynton, G. M., Ciaramitaro, V. M., & Arman, A. C. (2006). Effects of feature-based attention on the motion aftereffect at remote locations. *Vision Research*, *46*(18), 2968–2976. https://doi.org/10.1016/J.VISRES.2006.03.003

Brainard, D. H. (1997). The Psychophysics Toolbox Short Title: The Psychophysics Toolbox Corresponding Author. *Spatial Vision*, *10*, 433–436. Retrieved from http://color.psych.ucsb.edu/psychtoolbox

Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. (2000). Adaptive Rescaling Maximizes Information Transmission. *Neuron*, *26*(3), 695–702. https://doi.org/10.1016/S0896-6273(00)81205-2

Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research*, *38*(15–16), 2455–2468. https://doi.org/10.1016/S0042-6989(97)00469-0

Byatt, G., & Rhodes, G. (2004). Identification of own-race and other-race faces: Implications for the representation of race in face space. *Psychonomic Bulletin and Review*, *11*(4), 735–741. https://doi.org/10.3758/BF03196628

Campbell, J., & Burke, D. (2009). Evidence that identity-dependent and identity-independent neural populations are recruited in the perception of five basic emotional facial expressions. *Vision Research*, *49*(12), 1532–1540. https://doi.org/10.1016/J.VISRES.2009.03.009

Cavazos, J. G., Noyes, E., & O'Toole, A. J. (2019). Learning context and the other-race effect: Strategies for improving face recognition. *Vision Research*, *157*, 169–183. https://doi.org/10.1016/J.VISRES.2018.03.003

Clifford, C. W. G., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007a). Visual adaptation: Neural, psychological and computational aspects. *Vision Research*, *47*(25), 3125–3131. https://doi.org/10.1016/J.VISRES.2007.08.023

Clifford, C. W. G., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007b). Visual adaptation: Neural, psychological and computational aspects. *Vision Research*, *47*(25), 3125–3131. https://doi.org/10.1016/J.VISRES.2007.08.023

Collins, E., Behrmann, M., Tanaka, J. W., & C-N Wong, A. (2020). Exemplar learning reveals the representational origins of expert category perception. *Proceedings of the National Academy of Sciences*, *117*(20), 11167–11177. https://doi.org/10.1073/PNAS.1912734117

Dhruv, N. T., & Carandini, M. (2014a). Cascaded Effects of Spatial Adaptation in the Early Visual System. *Neuron*, *81*(3), 529–535. https://doi.org/10.1016/j.neuron.2013.11.025

Dhruv, N. T., & Carandini, M. (2014b). Cascaded Effects of Spatial Adaptation in the Early Visual System. *Neuron*, *81*(3), 529–535. https://doi.org/10.1016/J.NEURON.2013.11.025

Dosher, B. A., Jeter, P., Liu, J., & Lu, Z. L. (2013). An integrated reweighting theory of perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(33), 13678–13683. https://doi.org/10.1073/PNAS.1312552110/-/DCSUPPLEMENTAL/PNAS.201312552SI.PDF

Duan, X.-D., Wang, C.-R., Liu, X.-D., Li, Z.-J., Wu, J., & Zhang, H.-L. (2010). *"Ethnic Features Extraction and Recognition of Human Faces,."*

Elliott, S. L., Georgeson, M. A., & Webster, M. A. (2011). Response normalization and blur adaptation: Data and multi-scale model. *Journal of Vision*, *11*(2), 7–7. https://doi.org/10.1167/11.2.7

Engel, S. A., Wilkins, A. J., Mand, S., Helwig, N. E., & Allen, P. M. (2016). Habitual wearers of colored lenses adapt more rapidly to the color changes the lenses produce. *Vision Research*, *125*, 41–48. https://doi.org/10.1016/J.VISRES.2016.05.003

Fairhall, A. L., Lewen, G. D., Bialek, W., & de Ruyter van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature*, *412*(6849), 787–792. https://doi.org/10.1038/35090500

Fang, F., & He, S. (2005). Viewer-Centered Object Representation in the Human

Visual System Revealed by Viewpoint Aftereffects. *Neuron*, *45*(5), 793–800. https://doi.org/10.1016/J.NEURON.2005.01.037

Fang, F., Murray, S. O., Kersten, D., & He, S. (2005). Orientation-Tuned fMRI Adaptation in Human Visual Cortex. *Journal of Neurophysiology*, *94*(6), 4188–4195. https://doi.org/10.1152/jn.00378.2005

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, *20*(5), 453–467. https://doi.org/10.1037/h0059826

Haak, K. V., Fast, E., Bao, M., Lee, M., & Engel, S. A. (2014). Four days of visual contrast deprivation reveals limits of neuronal adaptation. *Current Biology*, *24*(21), 2575–2579. https://doi.org/10.1016/j.cub.2014.09.027

Haberman, J., & Whitney, D. (2009). Seeing the Mean: Ensemble Coding for Sets of Faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/A0013899

Hayward, W. G., Favelle, S. K., Oxner, M., Chu, M. H., & Lam, S. M. (2017). The other-race effect in face learning: Using naturalistic images to investigate face ethnicity effects in a learning paradigm. *Quarterly Journal of Experimental Psychology*, *70*(5), 890–896. https://doi.org/10.1080/17470218.2016.1146781

Htake, H., Tin, K., & Sein, M. M. (2011). Race Identification from Face Images Race Identification for Face Images. *ACEEE Int. J. on Information Technology*, *01*(02).

Jaquet, E., Rhodes, G., & Hayward, W. (2008). Race-contingent aftereffects suggest distinct perceptual norms for different race faces. *Visual Cognition*, *16*(6), 734–753. https://doi.org/10.1080/13506280701350647

Katti, H., & Arun, S. P. (2019). Are you from North or South India? A hard face-classification task reveals systematic representational differences between humans and machines. *Journal of Vision*, *19*(7), 1–1. https://doi.org/10.1167/19.7.1

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoobox-3? *Perception*, *36*(ECVP Abstract Supplement), 14.

Kohn, A. (2007a). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. https://doi.org/10.1152/jn.00086.2007

Kohn, A. (2007b). Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *J Neurophysiol*, *97*, 3155–3164. https://doi.org/10.1152/jn.00086.2007

Kovács, G., Zimmer, M., Bankó, É., Harza, I., Antal, A., & Vidnyánszky, Z. (2006). Electrophysiological Correlates of Visual Adaptation to Faces and Body Parts in Humans. *Cerebral Cortex*, *16*(5), 742–753. https://doi.org/10.1093/cercor/bhj020

Laughlin, S. (1981). *A Simple Coding Procedure Enhances a Neuron's Information Capacity*. *Z. Naturforsch* (Vol. 36).

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

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

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General*, *129*(4), 559. https://doi.org/10.1037/0096-3445.129.4.559

Li, Y., Tregillus, K. E., Luo, Q., & Engel, S. A. (2020). Visual mode switching learned through repeated adaptation to color. *ELife*, *9*. https://doi.org/10.7554/eLife.61179

Little, A. C., Debruine, L. M., & Jones, B. C. (n.d.). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. https://doi.org/10.1098/rspb.2005.3220

Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1578), 2283–2287. https://doi.org/10.1098/rspb.2005.3220

Little, A. C., DeBruine, L. M., Jones, B. C., & Waitt, C. (2008a). Category contingent aftereffects for faces of different races, ages and species. *Cognition*, *106*(3), 1537–1547. https://doi.org/10.1016/J.COGNITION.2007.06.008

Little, A. C., DeBruine, L. M., Jones, B. C., & Waitt, C. (2008b). Category contingent aftereffects for faces of different races, ages and species.

Cognition, *106*(3), 1537–1547. https://doi.org/10.1016/J.COGNITION.2007.06.008

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

McGovern, D. P., Roach, N. W., & Webb, B. S. (2012). Perceptual Learning Reconfigures the Effects of Visual Adaptation. *Journal of Neuroscience*, *32*(39), 13621–13629. https://doi.org/10.1523/JNEUROSCI.1363-12.2012

Meissner, C. A., & Brigham, J. C. (2001). Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review. *Psychology, Public Policy, and Law*. American Psychological Association Inc. https://doi.org/10.1037/1076-8971.7.1.3

Movshon, J. A., & Lennie, P. (1979). Pattern-selective adaptation in visual cortical neurones. *Nature*, *278*(5707), 850–852. https://doi.org/10.1038/278850a0

Olshausen, B. A., & Field, D. J. (1996). Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, *7*(2), 333–339. https://doi.org/10.1088/0954-898X_7_2_014

Patterson, C. A., Wissig, S. C., & Kohn, A. (2014). Adaptation Disrupts Motion Integration in the Primate Dorsal Stream. *Neuron*, *81*(3), 674–686. https://doi.org/10.1016/J.NEURON.2013.11.022

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

https://doi.org/10.1163/156856897X00366

Pestilli, F., Viera, G., & Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? *Journal of Vision*, *7*(7), 9–9. https://doi.org/10.1167/7.7.9

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. https://doi.org/10.1037/a0031568

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). *Practising orientation identification improves orientation coding in V1 neurons*. https://doi.org/https://doi.org/10.1038/35087601

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews. Neuroscience*, *8*(7), 522–535. https://doi.org/10.1038/nrn2155

Shapley, R., & Enroth-Cugell, C. (1984). Chapter 9 Visual adaptation and retinal gain controls. *Progress in Retinal Research*, *3*(C), 263–346. https://doi.org/10.1016/0278-4327(84)90011-7

Sigman, M., & Gilbert, C. D. (2000). *Learning to find a shape*. Retrieved from http://neurosci.nature.com

Soto, F. A., Escobar, K., & Salan, J. (2020). Adaptation aftereffects reveal how categorization training changes the encoding of face identity. *Journal of Vision*, *20*(10), 18. https://doi.org/10.1167/jov.20.10.18

Spivey, M. J., & Spirn, M. J. (2000). Selective visual attention modulates the direct tilt aftereffect. *Perception & Psychophysics 2000 62:8*, *62*(8), 1525–1533. https://doi.org/10.3758/BF03212153

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective, & Behavioral Neuroscience 2009 9:1*, *9*(1), 122–131. https://doi.org/10.3758/CABN.9.1.122

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Wainwright, M. J. (1999). Visual adaptation as optimal information transmission. *Vision Research*, *39*(23), 3960–3974. https://doi.org/10.1016/S0042-6989(99)00101-7

Wang, Y., Feng, Y., Liao, H., Luo, J., & Xu, X. (2018). Do They All Look the Same? Deciphering Chinese, Japanese and Koreans by Fine-Grained Deep Learning; Do They All Look the Same? Deciphering Chinese, Japanese and Koreans by Fine-Grained Deep Learning. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. https://doi.org/10.1109/MIPR.2018.00015

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

https://doi.org/10.1016/J.CONB.2007.07.001

Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science*, *1*(1), 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

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

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

Webster, M. A., & Leonard, D. (2008). Adaptation and perceptual norms in color vision. *Journal of the Optical Society of America A*, *25*(11), 2817. https://doi.org/10.1364/JOSAA.25.002817

Webster, M. A., & Maclin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review*, *6*(4), 647–653. https://doi.org/10.3758/BF03212974

Webster, M. A., & Mollon, J. D. (1995). Colour constancy influenced by contrast adaptation. *Nature*, *373*(6516), 694–698. https://doi.org/10.1038/373694a0

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research*, *37*(23), 3283–3298. https://doi.org/10.1016/S0042-6989(97)00125-9

Weibert, K., Harris, R. J., Mitchell, A., Byrne, H., Young, A. W., & Andrews, T. J. (2016). An image-invariant neural response to familiar faces in the human medial temporal lobe. *Cortex*, *84*, 34–42.

https://doi.org/10.1016/J.CORTEX.2016.08.014

Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the Cortical Hierarchy: Low-Level Curve Adaptation Affects High-Level Facial-Expression Judgments. *Journal of Neuroscience*, *28*(13), 3374–3383. https://doi.org/10.1523/JNEUROSCI.0182-08.2008

Yang, J., Yan, F. F., Chen, L., Xi, J., Fan, S., Zhang, P., … Huang, C. B. (2020). General learning ability in perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(32), 19092–19100. https://doi.org/10.1073/PNAS.2002903117

Yehezkel, O., Sagi, D., Sterkin, A., Belkin, M., & Polat, U. (2010a). Learning to adapt: Dynamics of readaptation to geometrical distortions. *Vision Research*, *50*(16), 1550–1558. https://doi.org/10.1016/j.visres.2010.05.014

Yehezkel, O., Sagi, D., Sterkin, A., Belkin, M., & Polat, U. (2010b). Learning to adapt: Dynamics of readaptation to geometrical distortions. *Vision Research*, *50*(16), 1550–1558. https://doi.org/10.1016/J.VISRES.2010.05.014