Norm- and Exemplar- Based Models of Face Recognition

David Andrew Ross

Thesis submitted to Cardiff University
for the degree of
Doctor of Philosophy

December, 2011

Page intentionally left blank

DECLARATION AND STATEMENTS

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ……………………………........(candidate) Date ……………………….

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ……………………………........(candidate) Date ……………………….

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ……………………………….....(candidate) Date ……………………….

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ……………………………….....(candidate) Date ……………………….

At the time of submission the data presented in Chapter 2 are published as:

Ross, D. A., Hancock, P. J. B., Lewis, M. B. (2010). Changing faces: Direction is important. *Visual Cognition*, 18, 67-81.

Acknowledgements

I would like to thank a number of people who have supported me during the time in which the work presented in this thesis was undertaken. In particular, thanks to Rob Honey and my good friends, Joe, Mikael, Mick and Damian all of whom helped minimise the stress in the tough times and make Cardiff an enjoyable place to live.

I would also like to thank my family, especially my father who has proof read enough of my work over the course of my BSc, MSc and PhD that he must surely be more than qualified to write a thesis of his own!

Thanks also to Isabel Gauthier, Thomas Palmeri, and Peter Hancock for providing valuable input to this work, it is difficult to imagine how different this thesis would be if it was not for their patient guidance and collaboration. Finally, a big thank you to my supervisors Michael Lewis and Dominic Dwyer for their input and patience at every stage of this thesis.

This work was funded by ESRC grant PTA-031-2006-00064

Table of Contents

List of Figures

Chapter 1

Literature Review

1.1 Introduction

Face recognition is vital for many of our day-to-day social interactions. For the most part we can effortlessly recognise the faces of family, friends, casual acquaintances and famous people, ignoring the new hairstyle or fashionable sunglasses that they might be sporting and attending to diagnostic features that reveal their identity. From a psychological or computational standpoint, our flair for face recognition is particularly interesting because, unlike the many other categories of object that we routinely encounter, faces must be individuated. That is to say, faces must be differentiated at the within-category level, placing a unique demand on the visual system's ability to rapidly and accurately discriminate a large number of visually similar patterns.

The ease with which we recognise familiar faces belies the computational complexity of the task, rendering us unaware of the dramatic image variance caused by changes in viewing angle, lighting conditions and partial occlusion, and leaving us with an illusion of stability that is characteristic of our impression of the visual world. Clearly, with some 120 million retinal cells semi-independently encoding the various aspects of a visual scene (Palmeri & Cottrell, 2010), the problem of face recognition is not (normally) one of visual acuity. Rather, the problem of face recognition is one of dimensional reduction, representing faces in a way that ignores the gross image-level variance across different instances of the same person while still maintaining the ability to discriminate between different faces (see Figure 1.1).

Figure 1.1. Unfamiliar face images taken from Burton and Jenkins (2011). **Left:** Two images of the same face taken under different conditions. **Right:** Two images of different faces taken under different conditions. Image level differences between two views of the same person can be dramatic.

Early approaches to face recognition, inspired by models in the object-recognition literature, worked on the assumption that the goal of vision was to reconstruct a faithful representation of the visual world (Edelman, 1999), identifying a limited set of non-accidental (view-invariant) object components, dubbed primitives (Marr, 1982) or geons (Biederman, 1987) that could be used to reconstruct a view-point and lighting invariant representation of each encountered object. Unfortunately, while influential, such theories appear to have important limitations. Primarily, with respect to face recognition, the extreme dimensional reduction and necessary loss of metric information required for extracting primitives or geons can result in a representation that is insufficient for making within-category discriminations (See Edelman, 1999 for a discussion). Moreover, it has become apparent that the extraction and identification of generic, non-accidental shape properties is far from being trivial.

An alternative approach, that is perhaps more suited to face recognition, is provided by image-based models of object recognition, in which objects are encoded by way of their similarity to previously encountered exemplars, storing representations of each encountered object across multiple view-points and lighting conditions[1] (e.g. Bulthoff & Edelman, 1992; Riesenhuber & Poggio, 1999). While some form of dimensional reduction still appears to be necessary, perhaps extracting dimensions that capture the greatest variance between a particular class of objects (e.g. Burton, Bruce, & Hancock, 1999), the power of the image-based approach comes from the fact that neural representations are

---

[1] The image-based framework may be divided into two strands of research. Some authors have suggested that mental representations of faces may make up a view-dependent structural code, whereas, others have suggested a pictorial code.

assumed to reflect the natural statistics of a given category of objects. Thus, rather than attempting to extract a faithful representation of the physical structure of each object, the main aim of visual recognition is identified as the representation of similarity (Edelman, 1999).

1.2 Face-Space Models of Face Recognition

Consistent with image-based approaches to object recognition, 'face-space' models of face recognition suggest that familiar and unfamiliar faces are represented in the visual system with respect to their similarity to the population of experienced faces, capturing the natural statistical variations along a set of diagnostic dimensions (e.g. Valentine, 1991). Newly encountered faces are compared to the stored representations of previously encountered faces in order to determine if the face is familiar or unfamiliar. Much of the literature on face-space models has focused on differentiating two possible representation schemes. On the one hand, faces may be encoded with respect to a prototype, or norm face, representing how each face deviates from the central tendency. On the other hand, faces may be encoded with respect to their similarity to the population of familiar face exemplars as a whole.

1.2.1 Norm-Based Models

The idea that faces may be represented with respect to a prototype, or norm face, has significant intuitive appeal, as evidenced by the central role that norms have taken in a number of theoretical accounts of face recognition (Giese &

Leopold, 2005; Goldstein & Chance, 1980; Loffler, Yourganov, Wilkinson, & Wilson, 2005; Rhodes, Brennan, & Carey, 1987; Rhodes & Jeffery, 2006; Valentine, 1991; Valentine & Bruce, 1986). With respect to face-space models, the norm is generally considered to be implicit, rather than explicit, as it is in prototype models of categorization, such that only the deviation of each face from this shared central point is encoded in memory (Valentine, 2001).

For example, Giese and Leopold (2005) provided a formal, mathematical, instantiation of the norm-based model (something that had been lacking in the literature) to account for neurophysiological data obtained from neurons in the macaque IT (Leopold, Bondar, & Giese, 2006). Consistent with previous verbal descriptions of the norm-based model (e.g. Loffler et al., 2005; Rhodes, 1996; Stevenage, 1997; Valentine, 1991), faces are represented by the direction and distance with which they deviate from the norm. Specifically, the activation of a face representation in response to a probe face is proportional to the cosine of the angle, relative to the norm, between the face representation and the probe. In addition, the distance of the probe face from the norm scales the overall activation such that faces further from the norm result in greater activation (see Chapter 4). Indeed, although the distance of a given probe face from the norm is represented, this information is only relevant for scaling the activation of the face representations. Accordingly, a probe face at any position along a trajectory from the norm, or 'identity trajectory' as it has been dubbed in the literature (e.g. Leopold, O'Toole, Vetter, & Blanz, 2001), will activate the same subsection of norm-coded face representations with only the overall level of activity scaled relative to the distance of the probe face from the norm (see Figure 1.2).

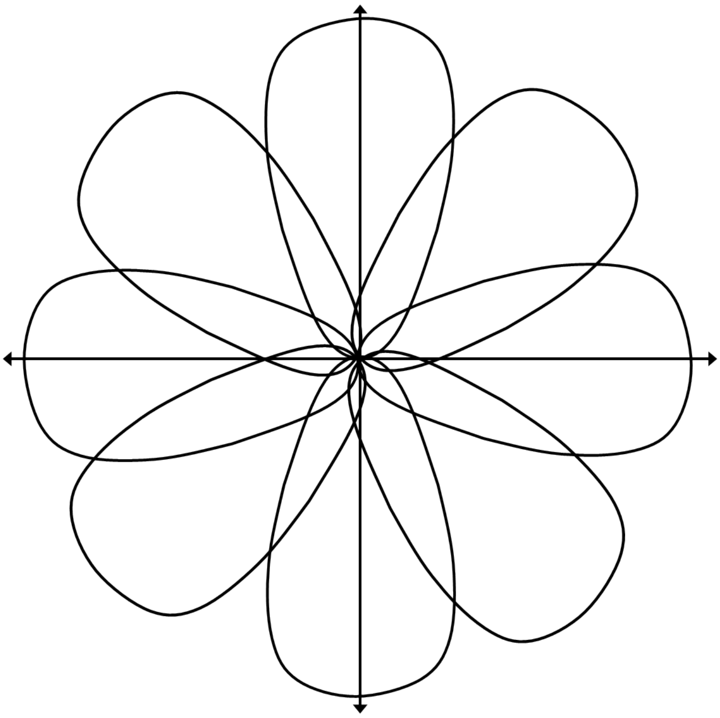


Figure 1.2. A representation of the neural activation in a norm-based face-space (e.g. Giese & Leopold, 2005). Each face representation, responds along a radial axis from the norm, or ‘identity trajectory’, with the activation strength, represented in the width of each ‘petal’, increasing for faces that are further from the norm. Importantly, a given representation only responds to deviations from the norm in a single direction, thus, crossing the norm will produce a negatively correlated pattern of activation.

It is often claimed that norm-based models offer an elegant strategy for encoding a population of faces, ignoring the shared variance between different faces in favour of encoding only the distinctive components of facial identity (Rhodes et al., 2005). However, the computational frugality of norm-based models comes at a cost; information about the overall distribution of faces in face-space is lost, as only the direction that each exemplar deviates from the norm is retained. Losing this information is not trivial, as male and female faces, the faces of different racial groups, different age groups, different families, as well as the faces of cartoon characters, actors, and even different species of animal, will all form distinct clusters in face-space. Consequently, in order to better capture the essential features of each cluster, it may be necessary to extend norm-based models, postulating multiple norms for at least the most densely populated of these clusters (Jaquet, Rhodes & Hayward, 2008; Valentine & Endo, 1992).

1.2.2 Exemplar-Based Models

In contrast to norm-based models, exemplar-based models do not specifically encode the position of the central tendency. Rather, faces are encoded by their similarity to all, or a subsection of, the other face exemplars. As a result, an advantage of exemplar-based models is that they directly reflect the overall, or local, statistics of a population of faces, with male and female faces, different race faces and other visually distinct groups represented by clusters in the face-space (Palmeri & Cottrell, 2010). Actually, as will be discussed in the next section, density information, represented within each cluster, is crucial to the predictions of exemplar-based models, in some cases providing a more parsimonious explanation for phenomena for which norm-based models must postulate multiple norms (e.g. Valentine & Endo, 1992).

Exemplar-based models of face recognition have generally made two assumptions. First, for a relatively homogeneous population of faces, such as those of the same race and gender, it is assumed that faces will be normally distributed along each of the face-space dimensions (Valentine, 1991). Given that the distribution of facial characteristics in a population represents natural variance, it is unsurprising that this assumption has received few challenges from studies that have measured the distribution of facial attributes (e.g. Burton, Bruce, & Dench, 1994). Second, for simplicity, and in the absence of evidence to the contrary, it is assumed that similarity is based on a Euclidean metric, such that the distance between a probe face and a given exemplar is the Euclidean distance between the two representations.

Unlike norm-based models, there have been a number of computational instantiations of the exemplar-based model. For example, Lewis (2004) instantiated a version, 'face-space-R', that can account for a number of phenomena that, on the surface, appear to be more consistent with a norm-based model. In Lewis' model, the face representations, or exemplars, are represented as locations, or vectors, in face-space, with the activation of each exemplar a Gaussian function of its similarity to the probe face (see Figure 1.3). The question of how broad a subsection of exemplars are activated in response to a given face is still in some debate with the precise way that exemplar activation contributes to recognition differing between models (e.g. Byatt & Rhodes, 1998; Lewis, 2004; Valentine, 1991). Nevertheless, density information is of central importance to all formulations of the exemplar-based model.

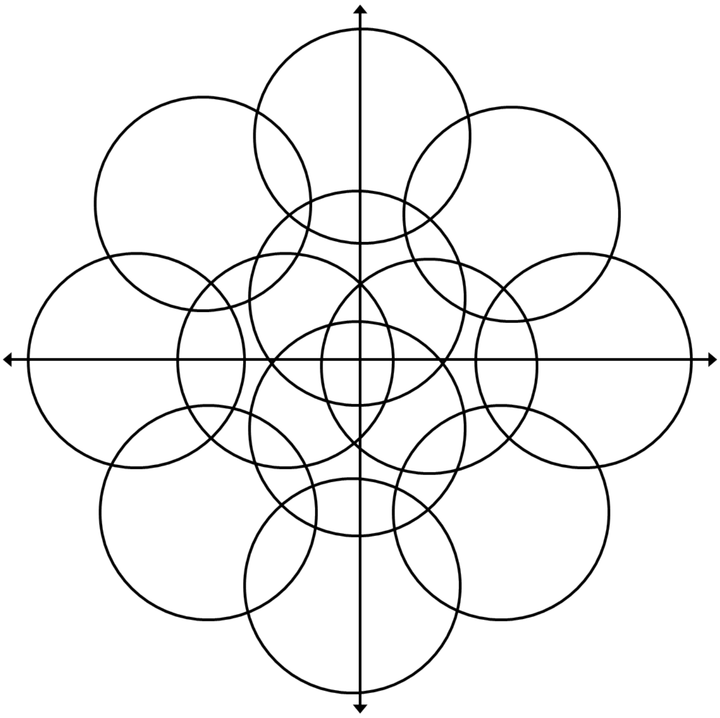


Figure 1.3. A representation of the neural activation in an exemplar-based face-space for a given population of faces. Each exemplar responds preferentially to a particular position in the face-space with the activation strength decreasing for targets further from this point.

1.3 Differentiating Norm- and Exemplar-Based Models

As has already been alluded to, a great deal of research effort has been devoted to differentiating norm- and exemplar-based models (e.g. Byatt & Rhodes, 1998; Giese & Leopold, 2005; Leopold, et al., 2001; Lewis, 2004; Lewis & Johnston, 1998; Loffler et al., 2005; Rhodes & Jeffery, 2006; Robbins, McKone, & Edwards, 2007; Valentine & Endo, 1992). However, much of this research has been unsuccessful in this respect, revealing, across a wide range of paradigms, that norm- and exemplar-based models can make very similar predictions. The remainder of Chapter 1 will provide a review of some of the main lines of research, describing the implications for norm- and exemplar-based face-space models.

1.3.1 Distinctiveness in Face Recognition

Distinctiveness is a salient feature of our psychological representation of faces, even in everyday conversation we might talk about facial distinctiveness when describing someone's appearance (see Figure 1.4). Within the face-space framework, distinctive faces are defined as faces that lie far from the centre of the face-space with typical faces lying closer to the centre (Johnston, Milne, Williams, & Hosie, 1997; Valentine, 1991). As such, typical faces will be closer to the internally represented norm, in a norm-based model, or closer to other face representations, in an exemplar-based model. Indeed, part of the original motivation for face-space was to provide a parsimonious account of the effects of distinctiveness in face recognition.

Figure 1.4. Typical and distinctive faces from a set of 200 as rated by participants (see Frowd et al., 2006). **Top:** The three faces that were rated as being most typical. **Bottom:** The three faces that were rated as being most distinctive.

For example, Valentine and Bruce (1986) reported that typical faces are categorised, as faces, more rapidly than distinctive faces; a finding that was initially understood in terms of a norm-based model, with typical faces sharing more in common with the norm than more distinctive faces. However, it turns out that this finding is also a direct prediction of exemplar-based models. Specifically, typical faces are closer to and, hence, activate more strongly, a greater number of other face exemplars than distinctive faces, which lie further from the centre of the space and, thus, activate fewer other exemplars (Valentine, 1991; also see Nosofsky, 1988). In other words, the ease of categorisation is determined by a 'summed similarity rule', whereby, typical faces have a greater summed similarity to all other face exemplars than more distinctive faces.

Additionally, and somewhat paradoxically, it turns out that when it comes to face recognition, as opposed to face categorisation, it is distinctive faces that are recognised faster (Valentine & Bruce, 1986), and remembered better (e.g. Light, Kayra-Stuart, & Hollander, 1979) than typical faces. Again, although this finding was initially understood in terms of comparison to a norm, with distinctive faces conveying more unique identity information (i.e. producing more activation in the corresponding face representation), it is also a direct prediction of an exemplar-based model. Specifically, Valentine (1991) suggested that the distinctive face recognition advantage ought to emerge in an exemplar-based model because, as a result of the relative scarcity of exemplars in the outer regions of face-space, distinctive faces will activate fewer competing exemplars. In other words, whereas categorisation is based on the similarity of a given probe to all members of the category, recognition is based on the activation of a given exemplar with its similarity to all other exemplars as a denominator (e.g. e.g. Nosofsky, 1988; Valentine, 1991).

While on the topic of distinctiveness it is worth noting that, in spite of the fact that faces are assumed to be normally distributed along each of the dimensions of face-space, it does not follow that most faces will be very typical (Burton and Vokey, 1998). Indeed, participant ratings of facial distinctiveness suggest that few faces are very distinctive or very typical, rather, most faces lie somewhere between these two extremes (Wickham, Morris, & Fritz, 2000). In fact, Burton and Vokey showed that this pattern of results would emerge as a natural consequence of a high dimensional representational space, with the faces in very high dimensional spaces tending to be distributed in a narrow region some distance out from the centre of the space. However, while important for understanding distinctiveness ratings, this does not change the fact that, in an exemplar–based model, faces will be more densely clustered at the centre of face-space (i.e. typical face exemplars are closer to other exemplars than distinctive exemplars).

1.3.2 The Caricature Effect

Research into the effect of caricature on face recognition has been central to the development of the norm-versus-exemplar debate (see Rhodes, 1996 for a review). Part of the reason for the interest in caricatures is the observation that they seem to be 'super portraits', somehow capturing the identity of the person being caricatured better than a faithful portrait or photograph. Indeed, a number of studies indicate that caricatures are recognised more quickly (Benson & Perrett, 1994; Rhodes et al., 1987) and more accurately (Benson & Perrett, 1994;

Rhodes, Carey, Byatt, & Proffitt, 1998; Rhodes & Tremewan, 1994) than the veridical images from which they were created.

Importantly, unlike the artist drawn caricatures of famous political and celebrity figures, typically seen in newspapers and magazines, the caricatures used in psychological research are simply literal exaggerations of the features of a face, as well as their configuration, relative to the population average. Roughly speaking, caricatures are generated by identifying corresponding points on both an average face and target face, calculating the difference between these points and then either exaggerating the difference, to create a caricature, or attenuating the difference, to create an 'anti-caricature' (e.g. Benson & Perrett, 1991a; Brennan, 1985; see Figure 1.5). It should be noted that, if anything, the effect of anti-caricature on face recognition is even more dramatic than the effect of caricature, with anti-caricature causing faces to be recognised considerably more slowly and less accurately than the veridical images from which they were created[2] (Benson & Perrett, 1994; Rhodes et al., 1998; Rhodes et al., 1987).

Seen in these terms, the relevance of caricature effects to the norm-versus-exemplar debate is clear; caricature exaggerates precisely those features that norm-based models suggest should be important to face recognition. That is to say, while caricaturing does not change the direction that a given face deviates from the norm it does make it more distinctive. In terms of a model such as Giese and Leopold's (2005) norm-based model, this would mean that the same face representations would be active for both the veridical and the caricature but in

---

[2] The fact that anti-caricature hinders face recognition is not, in its self, surprising. Indeed, this is what one would expect when distorting the appearance of a familiar face. However, what is theoretically relevant is the relationship between caricature and anti-caricature. Namely, that given an equal caricature and anti-caricature distortion, the caricature will be much better recognised than the anti-caricature.

the case of the caricature the representations would activate more strongly (presumably resulting in improved recognition).

Perhaps somewhat less clear is that exemplar-based models can also predict the caricature effect (e.g. Lewis, 2004; Lewis & Johnston, 1998; Lewis & Johnston, 1999a). Specifically, in an exemplar-based model, the caricature effect emerges as a result of the exemplar density gradient between the centre of the face-space and its outer reaches. For example, in the face-space-R model, the presentation of a familiar probe face results in all exemplars in the space becoming active to some degree (based on a Gaussian function of the Euclidean distance), with the exemplars competing in a winner-takes-all fashion to determine which face, if any, the probe is identified as. With respect to caricature, what is important about face-space-R is that the final recognition is not simply dependent on the activation of the target face exemplar but also on the activation of all other exemplars in the space. Lewis demonstrated that while a caricatured version of a familiar face will be further from the corresponding exemplar representation, it will activate the target face exemplar proportionally more than the veridical face will (also see Lewis & Johnston, 1999a; for a slightly different account).

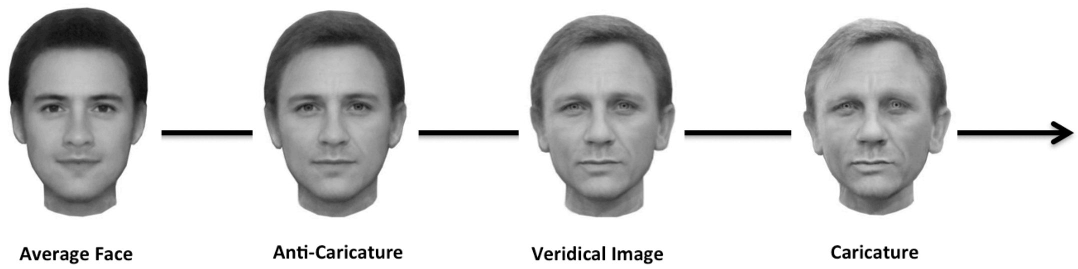


Figure 1.5. Computer generated caricatures of a famous face (Daniel Craig). From left to right with the black arrow representing increasing distinctiveness: An average face, an anti-caricature, the veridical image, and a caricature. Importantly, although the anti-caricature and the caricature are in fact the same distance from the veridical image, moved 50% closer to the average and 50% further from the average respectively, the caricature is generally considered to be a better likeness.

Indeed, while it might appear that both norm- and exemplar-based models are equally capable of predicting caricature effects it is actually possible that exemplar-based models provide a better account of the precise pattern of results. For example, in an exemplar-based model the beneficial effects of caricature will generally be restricted to relatively small degrees of caricature, at least for low or medium dimensional spaces (Lewis, 2004). Indeed, while an exemplar-based model would always predict that caricatures should be easier to recognise than the corresponding anti-caricatures, the actual degree of caricature that will result in improved recognition is more limited. In line with these predictions Benson & Perrett (1991b) estimated the size of the caricature effect to be in the region of 4.4%, an estimate that Lewis suggests is in line with a 15 to 20 dimensional face-space. In contrast, it would seem that norm-based models ought to predict that even very high degrees of caricature should be beneficial to recognition, since each identity is represented by its direction of deviation from the norm, or identity trajectory (Leopold et al., 2001). Thus, up to the point where the caricatured face ceases to be face-like, one might expect caricature to benefit recognition.

Moreover, the exemplar-based model predicts that caricature will be most effective if it is done relative to the corresponding race average (e.g. Byatt & Rhodes, 1998) a prediction that emerges naturally from the clustering of exemplars in face-space. Indeed, in order to account for these findings it would be necessary for a norm-based model to postulate the existence of multiple face norms, which, while possible, would seem to lack parsimony. Furthermore, Lewis (2004, also see Lewis & Johnston, 1999a) demonstrated that, in line with the empirical data, typical faces (Benson & Perrett, 1994) and degraded images, such as line drawings (Rhodes et al., 1987), would benefit from larger degrees of caricature. In contrast, norm-based models, which are still frequently cited as providing a natural account of caricature (e.g. Leopold et al., 2001; Rhodes et al., 2005; Rhodes & Jeffery, 2006; Tsao & Freiwald, 2006), have not yet been systematically tested on any of the above predictions.

1.3.3 Identity Trajectories

In order to determine what it is that endows caricatures with their psychological status as 'super portraits' it has proved useful to consider two alternative forms of caricature. As described in the previous section, the norm-based model suggests that both caricature and anti-caricature distort a face along its identity trajectory, emanating from the norm. However, it is also possible to distort a face in a direction that is lateral (Rhodes, 1996; Rhodes et al., 1998), or oblique (Lewis & Johnston, 1998), to the identity trajectory. Importantly, as Rhodes pointed out, if the identity trajectory really does represent a psychologically privileged direction in face-space then lateral or oblique caricatures ought not to be very well recognised. Indeed, Rhodes suggested that they ought to be recognised less accurately than anti-caricatures.

However, testing has revealed that it is actually anti-caricatures that are the hardest to recognise (Lewis & Johnston, 1998; Rhodes et al., 1998), followed by lateral or oblique caricatures, with standard caricatures recognised best. Notably, this pattern of results turns out to be precisely what is predicted by an exemplar-based account (Lewis & Johnston 1998; Lewis & Johnston, 1999a). That is to say, while caricature moves a face into a region of lower exemplar density and anti-caricature moves it to a region of higher exemplar density, lateral or oblique caricatures do something in between. That is to say the move the face into a region of intermediate exemplar density. Additionally, Lewis & Johnston (1999b) were actually able to demonstrate a causal link between the preferred direction of caricature and the presence of other competing exemplars. Lewis & Johnston taught participants to recognise two very similar faces. Next, they had them engage in a task where they had to identify one of the two faces from a set of faces morphed along the trajectory between and beyond the two faces. Their results indicated that the preferred morph was generally one that was shifted slightly away from the previously learned competing exemplar (as opposed to an exaggeration relative to the average face as predicted by the norm-based model).

Another prediction of the identity trajectory hypothesis, that is to say, the hypothesis that faces are represented by identity trajectories, which has received little empirical validation, is that the transition across the norm ought to reflect a discontinuity in the perception of identity (e.g. Blanz, O'Toole, Vetter, & Wild, 2000; Leopold et al., 2001; Rhodes et al., 2005). Indeed, if direction is important to the representation of faces then pairs on opposite sides of the norm should genuinely be perceived as opposite (at this point in face-space the angle between two faces will be at its greatest). However, it seems that, if anything, the perceived similarity of face-pairs that traverse the norm is actually greater than it is for face-pairs that lie on the same trajectory but that do not traverse the norm (Rhodes, Maloney, Turner, & Ewing, 2007, though see Blanz et al., 2000). (Note, this research, as well as the research on lateral and oblique caricatures, will be returned to in more detail in Chapters 2 and 3).

1.3.4 Interim Summary

In summary, it would seem that, contrary to some recent claims (Tsao & Freiwald, 2006; Rhodes et al., 2005; Leopold et al., 2001), norm-based models do not offer a particularly compelling account of the effects of caricature on face recognition - at least, not when they are contrasted with exemplar-based accounts. Indeed, it could be argued that exemplar-based models might actually provide a better account of the overall findings. There are, however, some important caveats to this conclusion.

One such caveat is that, while exemplar-based models have been formalised and their predictions tested, the same cannot be said for norm-based models (with the exception of Giese & Leopold, 2005). Considering the clear parallels that could be drawn between the present debate and the exemplar-versus-prototype debate in the categorisation literature, it is surprising that formal modelling has not played a more prominent role. By comparison, the categorisation literature has featured extensive research on the relation between exemplar theory and prototype theory, leading to a deep understanding of the assumptions that these models entail (Pothos & Wills, 2011). Yet, in face recognition it is left up to the intuitions of the author to determine how a particular model might behave.

A second, and possibly related, caveat is that, descriptions of the norm-based model vary from author-to-author. Up until now only versions of the norm-based model that posit direction to be important have been considered (e.g. Giese & Leopold, 2005; Loffler et al., 2005; Rhodes, 1996; Valentine, 1991). These models are the most straightforward interpretation of the norm-based model. However, there have also been descriptions of the norm-based model in which direction is unimportant to the similarity metric (e.g. Craw, 1995; Meytlis, 2011). In these versions of the norm-based model it would appear that the norm is merely assumed to be a reference point for the location of each face exemplar, an assumption that would mean that it plays no role in the actual similarity metric of face-space. As a result these versions of the norm-based model can reasonably be assumed to make the same predictions as an exemplar-based model.

For now it is sufficient to conclude that there is little empirical evidence that direction (relative to the norm) is important to the psychological representation of faces. However, it would be wrong to think that this rules out a norm-based model of face recognition for the two reasons discussed above. That is, first, norm-based models have not been sufficiently explored through computational modelling and, second, some versions of the norm-based model do not postulate that direction is important. In fact, as will be discussed in the following section, it turns out that there is actually a great deal of empirical support for the idea that faces are encoded with respect to a norm. It is worth highlighting here that, for the remainder of the thesis, norm-based models that assume direction to be important will be referred to as *traditional norm-based models*.

1.4 Adaptation Aftereffects

An important property of face-space models is that they represent faces with respect to past experience. The face-space must be updated, perhaps changing the location of the norm, or adding a new exemplar representation, in response to newly encountered faces. This somewhat understates the experimentally revealed situation; it has recently become apparent that face-space is even more dynamic than was previously considered, with even a brief period of exposure to a given face (less than a second) sufficient to cause a measurable bias, or aftereffect, in the perception of a subsequently presented face (Rhodes, Jeffery, Clifford, & Leopold, 2007). For example, adapting to a face that has had its internal features expanded, as if the image had been projected onto a convex surface, biases the perception of a subsequently presented average-face such that the features appear contracted, as if projected onto a concave surface (Webster & MacLin, 1999). This section will review the growing body of literature that has attempted to use the short-term adaptive properties of face-space to probe the nature of face representations.

Leopold et al. (2001) have demonstrated that face aftereffects can be used to tap into the neural mechanisms that encode facial identity. To do this they created morph trajectories, extending from four target identities (Adam, Jim, John and Henry) passing through an average face and beyond, to create four 'anti-faces' on the opposite side of the average-face. Leopold et al. found that the identification of a face on the continuum between the average face and the target face, Adam, was facilitated by a brief period of adaptation to the corresponding anti-face, anti-Adam, but hindered by adaptation to one of the three non-corresponding anti-faces (see Figure 1.6), suggesting that exposure to the anti-face had biased perception towards the opposite target face.

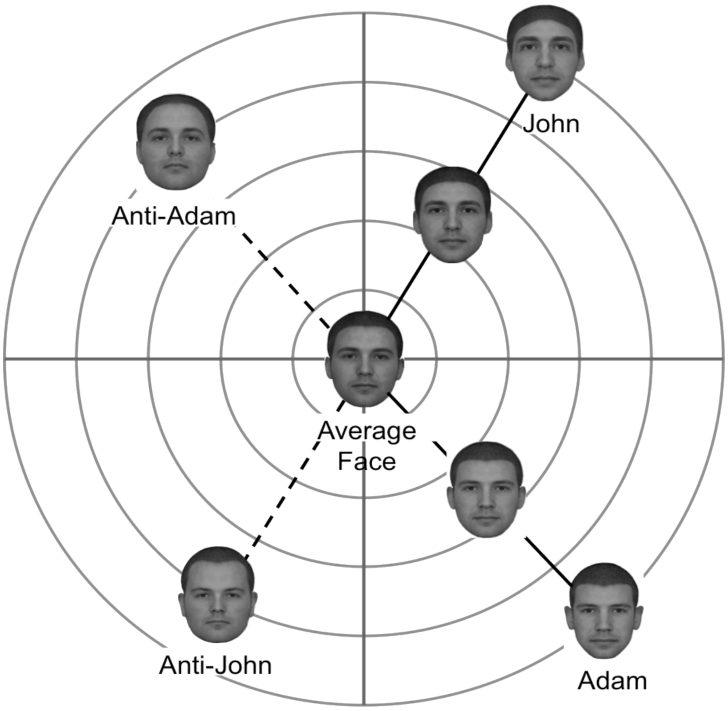


Figure 1.6. Schematic representation of the stimuli used by Leopold et al. (2001). All identity trajectories (shown here for two faces, Adam and John) pass through the average face such that the 0% identity level of each target trajectory is equivalent to the average-face. Anti-faces lay -80% opposite each of the corresponding faces.

Importantly, face adaptation is invariant in response to changes in simple stimulus properties such as size (e.g. Zhao & Chubb, 2001), colour (e.g. Yamashita, Hardy, De Valois, & Webster, 2005), and retinal position (e.g. Leopold et al., 2001; Fang & He, 2005) that are thought to be encoded in low-level vision, suggesting that it does indeed tap into more complex representations such as those used to encode faces. Furthermore, face aftereffects are larger for face attributes that have a greater natural variability (Hills, Holland, & Lewis; 2010; Robbins, et al., 2007), strongly suggesting that they can be used to explore the differences in the type of statistical information used to encode facial shape in norm- and exemplar-based face-space models.

1.4.1 Adaptation in Norm- and Exemplar-Based Models

Exemplar-based accounts of face adaptation generally explain aftereffects in terms of a reduction in the activation-potential of face exemplars in proportion to their degree of activation to the adaptor face (e.g. Hurlbert, 2001; Rhodes & Jeffery, 2006; Robbins, et al., 2007; Susilo, McKone, & Edwards, 2010a). Thus, because activation in an exemplar-based model is based on the Euclidean distance between a probe face representation and each of the exemplar-representations, adaptation will be centred on the face-space location of the adaptor face (see Figure 1.7). If, subsequent to adaptation, a new face is presented, the exemplars in the adapted region of the face-space will respond less strongly than usual causing the perception of the new face to be biased away from the adaptor location. Thus, adaptation in an exemplar-based model can be thought of in terms of a general bias away from the adaptor location.

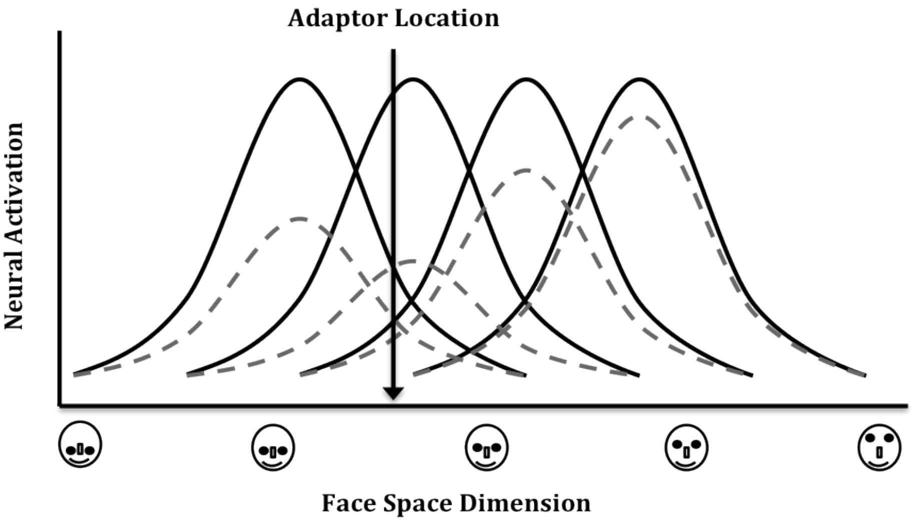


Figure 1.7. Simplified representation of an exemplar-based model along a single arbitrary dimension representing eye height. Solid lines represent potential neural responses of exemplar representations (Gaussian functions). Dotted lines represent the response potential following adaptation to a face at the location indicated. The most adapted exemplars will be the ones closest to the location of the adaptor.

In contrast, norm-based accounts of face adaptation generally explain aftereffects in terms of a change in the location of the norm relative to which all faces are encoded (e.g. Leopold et al., 2001; Rhodes & Jeffery, 2006; Robbins et al., 2007). Specifically, adaptation in a norm-based model occurs as a result of the norm being temporarily shifted in the direction of the adaptor face with the magnitude of the aftereffect dependent on the position of the adaptor relative to the norm (see Section 1.5 for a more mechanistic account). Specifically, adaptation in a norm-based model causes a bias in perception that is opposite to the way that the adaptor deviates from the norm. Thus, if the adaptor face has a particularly large nose then (as a result of the norm being moved in the direction of the adaptor) a subsequently presented target face will be perceived as having a smaller nose than it otherwise would.

As, a number of authors have pointed out (e.g. Rhodes et al., 2005; Rhodes & Jeffery, 2006; Tsao & Freiwald, 2006) both norm- and the exemplar-based models of adaptation are able to explain the findings reported by Leopold et al. (2001). Specifically, the facilitatory effect that adaptation to an anti-face, anti-Adam, has on the identification of the target face, Adam, could either be explained as a general bias away from the adaptor location (exemplar account) or as a bias in perception towards the opposite face (norm account). In both cases adaptation would cause targets on the morph trajectory between the average face and the target face, Adam, to be biased towards the target face.

Rhodes & Jeffery (2006) suggested that, in order to differentiate the 'general bias' predicted by an exemplar-based model from 'opposite bias' predicted by the norm-base model, it is necessary to introduce non-opposite morph trajectories into the paradigm used by Leopold et al. (2001). Thus, whereas the opposite trajectories extended from each of the four target faces through the average face to the anti-faces on the other side, the non-opposite morph trajectories were the same length as the opposite trajectories but did not pass through the average face (see Figure 1.8). They reasoned that, if adaptation resulted in a general bias away from the adaptor location then adaptation to the non-opposite adaptor face ought to facilitate the identification of morphs on the non-opposite trajectory just as much as adaptation to the opposite face facilitated the recognition of morphs on the opposite trajectory. However, if adaptation resulted in a bias towards an opposite face then adaptation to the non-opposite adaptor should facilitate identification substantially less than adaption along the opposite trajectory. Indeed, this is what they reported. That is to say, there was substantially less adaptation along the non-opposite trajectory than along the opposite-trajectory.

Also supporting the norm-based model of adaptation, Leopold and Bondar (2005) demonstrated that when the anti-face, used in the Leopold et al. (2001) paradigm, is exchanged for an average face the facilitatory effect of adaptation on the identification of the target face is almost eliminated. This finding would be consistent with the idea that more distinctive faces should cause a bigger change in the perceived location of the norm, resulting in a larger adaptation aftereffect. Moreover, Susilo, McKone, and Edwards (2010b) demonstrated that the magnitude of face aftereffects increases linearly as the adaptor is moved further from the average face (also see Susilo, et al., 2010a; Webster & MacLin, 1999).

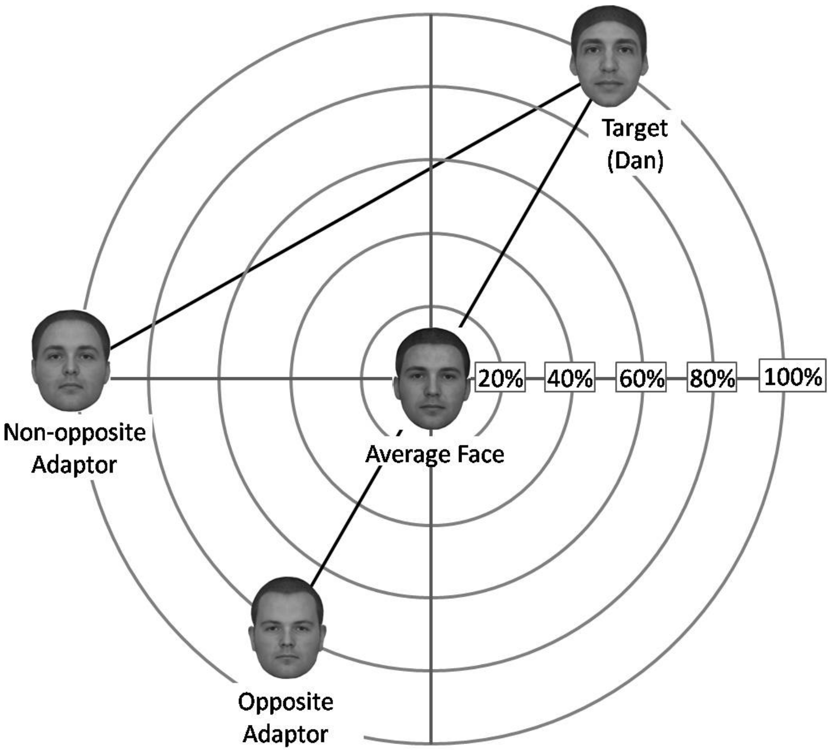


Figure 1.8. Schematic representation of the stimuli used by Rhodes et al. (2006). Opposite-trajectories (shown here extending between the opposite-adaptor and the target face, Dan) passed through the average face. Non-opposite trajectories (shown here extending between the non-opposite adaptor and the target face, Dan) were the same length as opposite trajectories but did not pass through the average face.

Finally, Robbins et al. (2007) provided fairly conclusive evidence that adaptation cannot simply be considered as a general bias away from the adaptor location. In their study, they created a continuum of faces between a face with both eyes raised as far as the hairline (raised 50 pixels) and an undistorted target face (eyes raised 0 pixels). They then asked participants to rate how similar faces at each point along the continuum were to the target face both before and after adaptation to a face at an intermediate point along the continuum (eyes raised 20 pixels). They reasoned that, if the face aftereffect was a general bias away from the adaptor location then, following adaptation, faces that were more extreme than the adaptor (eyes raised more than 20 pixels) ought to appear even less like the target than they did before adaptation. In contrast, faces that were less extreme than the adaptor (eyes raised by less than 20 pixels) ought to appear more similar to the target. However, this is not what they found. Rather, they found that, in line with a norm-based account of adaptation, both the more extreme and the less extreme faces were rated as more similar to the target following adaptation. Thus, adaptation appeared to have shifted the perception of the faces towards a face with opposite attributes to the adaptor.

In summary, it would seem that adaptation aftereffects provide a useful tool for exploring the nature of face representations in the visual system. Indeed, norm- and exemplar-based models appear to make different predictions across a range of adaptation paradigms with the results widely interpreted as favouring a norm-based account (e.g. Leopold & Bondar, 2005; Leopold et al., 2001; Rhodes & Jeffery, 2006; Rhodes et al., 2005; Robbins, et al., 2007; Susilo, et al., 2010a; Susilo, et al., 2010b; see Rhodes & Leopold, 2011 for a recent review).

1.5 The Two-Pool Norm-Based Model

The above description of adaptation in a norm-based model might be considered to be somewhat vague. The norm is 'updated' or 'moved' following exposure to an adaptor face but in comparison to descriptions of exemplar-based adaptation it is less clear how this might be implemented. This has led some researchers to propose a new account of the norm-based model, one that provides an explicit mechanism behind the changing location of the norm, the two-pool model (Rhodes & Jeffery, 2006; Robbins et al., 2007). The two-pool model is based on models from low- and mid-level vision (e.g. Mather, 1980; Regan & Hamstra, 1992) that have been used to describe qualitatively similar aftereffects in motion perception and aspect ratio. In general, it appears to provide a clear description of the data observed in studies of face adaptation.

In this type of model, face representations are encoded such that the location of a face along a given dimension is represented by the relative activity of two opponent-coded neural pools (see Figure 1.9). As a result, the location of the norm is implicit in the opponent coding such that the presentation of an average-face would result in equal activation of both pools. If, instead of an average-face, a distinctive face is presented, say a face with eyes raised up, as in the study by Robbins et al. (2007), this face will now activate one pool more strongly than the other. This feature is central to the way that the two-pool model accounts for the face aftereffects reviewed in Section 1.4

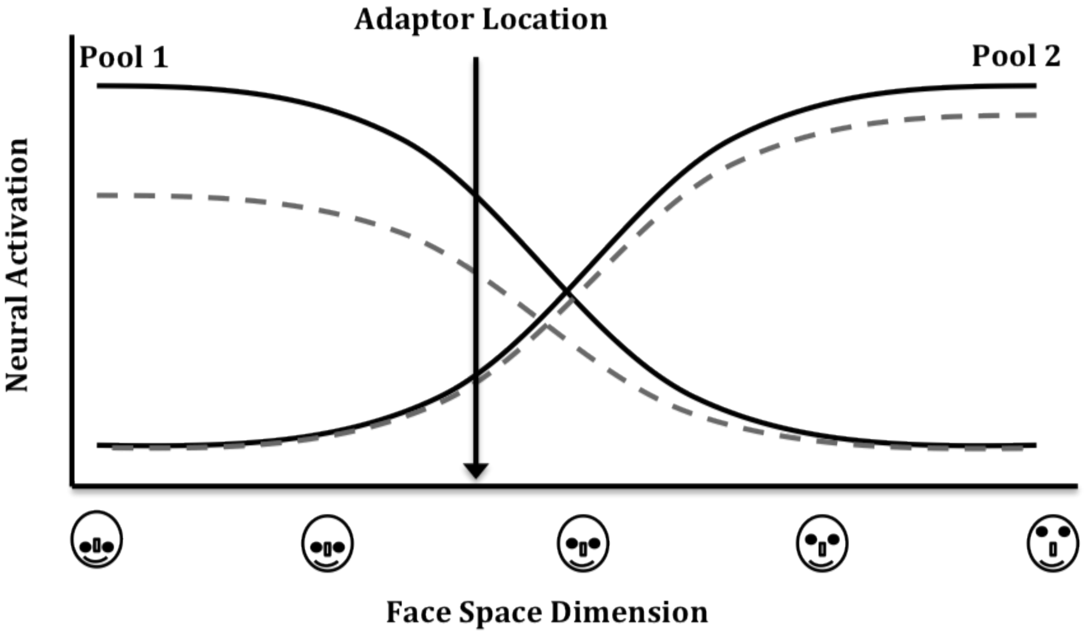


Figure 1.9. Simplified representation of a two-pool model along a single arbitrary dimension. Solid lines represent potential opponent neural responses. Dotted lines represent the response potential following adaptation to a face at the location indicated. Adaptation will cause a bias in the perception of a face towards a face with opposite attributes to the adaptor.

Specifically, descriptions of the two-pool model assume that adaptation will cause a reduction in the firing of a given pool in proportion to its degree of activation. As a result, exposure to an average face will not affect subsequent perception. Rather, the opponent-coded pools will be activated equally resulting in no change in their proportional activation. In contrast, if a distinctive face is presented it will activate one pool more strongly than the other, causing its response to become proportionally weaker, biasing subsequent perception in the opposite direction. Thus, the two-pool model would appear to be consistent with the observed face aftereffects, predicting that adaptation will bias perception towards an opposite face, with the magnitude of the aftereffect increasing as the adaptor is moved further from the norm.

The relationship between the two-pool model and previous versions of norm-based face-space (e.g. Giese & Leopold, 2005; Valentine, 1991) is somewhat unclear. Rhodes et al. (2005) suggested that each pair of opponent-pools would encode the location of a face down a given dimension (with the pools aligned with the dimensions) and specifically rejected the interpretation that the pools would represent faces and their opposing anti-faces (e.g. Hurlbert, 2001). However, if the pools are not assumed to correspond in some way to face representations then it is unclear that the two-pool model is really a face-space model.

For these reasons, there appear to be two possibilities for instantiating the two-pool model. On the one hand, one could implement a model with the face-space dimensions explicitly represented by opponent coded pools centred on the norm and then add a second 'face-space' layer in which exemplars are coded as normal. This would seem to be in line with the descriptions given by

Rhodes et al. (2005), though it is unclear why they favour this account. On the other hand, one could implement a model in which face representations were explicitly encoded in opposition to an anti-face, removing the need for a second layer (a model such as this is implemented in Chapter 4). Indeed, how ‘different’ these two approaches are may come down to how literally one takes the face-space metaphor. Thus, for the remainder of the thesis, *two-pool model* will refer to the general class of norm-based models that assume some form of norm-centred opponent coding.

1.6 Overview of the Current Work

Norm- and exemplar-based models of face-space have provided the basis for rich debate within the face recognition literature. As discussed, recent research into face adaptation has led to an overwhelming shift in the field towards some form of norm-based model. Indeed, this shift is evidenced by the slew of recent research articles claiming to find evidence of norm-based adaptive coding in both adults (Freiwald, Tsao, & Livingstone, 2009; Griffin, McOwan, & Johnston, 2011; Leopold & Bondar, 2005; Leopold et al., 2001; Rhodes, et al., 2011; Rhodes & Jeffery, 2006; Rhodes, Watson, Jeffery, & Clifford, 2010; Robbins, et al., 2007; Susilo, et al., 2010a; Susilo, et al. 2010b; Tsao & Freiwald, 2006), and children (Anzures, Mondloch, & Lackner, 2009; Jeffery et al., 2010; Nishimura, Maurer, Jeffery, Pellicano, & Rhodes, 2008; Nishimura, Robertson, & Maurer, 2011; Pimperton, Pellicano, Jeffery, & Rhodes, 2009; Short, Hatry, Mondloch, 2011).

Nonetheless, questions still remain about the nature of norm-based coding. For example, as was discussed in Sections 1.3.2 and 1.3.3 a number of descriptions of the norm-based model seem to suggest that facial identity is represented as a direction of deviation from the norm, a position that is apparent in many verbal descriptions of caricature throughout the literature (e.g. Blanz et al., 2000; Leopold et al., 2001; Rhodes et al., 2005). However, evidence for these predictions has not been forthcoming. Importantly, while the results of studies into face adaptation do seem to support the idea that the norm plays a role in the encoding of faces, they have not yet been able to address whether direction is important. Thus, Chapters 2 and 3 will revisit the research into lateral and oblique caricature in order to examine the role of direction, relative to the norm, in the representation of faces.

A second major issue that was identified in the present literature review is that face-space models, in particular norm-based models, have rarely been tested through computational implementation. Indeed, given that face-space models contain several properties that are notoriously difficult to intuit, such as high dimensionality and non-linearity (Hintzman, 1990), it would seem crucial that formal comparisons between these two approaches are carried out. Thus, in Chapter 4 three versions of face-space, an exemplar-based model, a traditional norm-based model (e.g. Giese & Leopold, 2005) and a two-pool norm-based model (e.g. Rhodes & Jeffery, 2006) were implemented, and their predictions tested on the full range of putatively diagnostic adaptation paradigms.

Finally, In Chapter 5 three studies looking into adaptation in novel objects are reported. Though preliminary, the aim of these studies was to investigate the extent to which long-term exposure to a stimulus set is necessary for the emergence of the type of aftereffects observed in studies of face recognition.

Chapter 2

Discriminating Caricature and Oblique-Caricature Face Pairs

Abstract

Chapter 2 reports two experiments investigating whether discrimination between faces lying on a caricature trajectory differs from discrimination between faces lying on an oblique trajectory. Experiment 1 tested discrimination thresholds for upright familiar and unfamiliar face-pairs, comparing the discrimination thresholds for face-pairs that lay on a caricature-trajectory with discrimination thresholds for pairs that lay on an oblique-trajectory. In line with traditional accounts of the norm-based model, in which the direction of deviation from the norm is psychologically important, the mean discrimination threshold was significantly higher for Caricature-Trajectory pairs than for the Oblique-Trajectory pairs. In Experiment 2 discrimination thresholds were tested for inverted familiar and unfamiliar face-pairs. There was no significant effect of trajectory type (Caricature- or Oblique-Trajectory) on discrimination threshold for inverted faces. However, a combined analysis of Experiments 1 and 2 revealed that, while there was an overall main effect of trajectory type, there was no interaction between trajectory type and orientation (upright or inverted). The results are taken as providing some support for the idea that direction, relative to the norm, is important in the representation of faces.

2.1 Introduction

As discussed in Chapter 1, norm-based face-space models have recently found favour in the literature. This favourable treatment is supported by research into the effects of face adaptation (e.g. Leopold et al., 2001; Rhodes et al., 2005; Robbins, et al., 2007). In spite of this, although face adaptation has provided evidence that faces are encoded with respect to a norm, the precise role of the norm in the psychological representation of faces is still unclear. On the one hand, faces might be encoded with respect to their direction of deviation from the norm – as in traditional accounts of the norm-based model (e.g. Tsao & Freiwald, 2006; Rhodes et al., 2005). On the other hand, the norm may merely serve as a reference point, with face adaptation realigning the axes, or dimensions, of face-space (e.g. Craw, 1995; Valentine, 2001).

In principle, it seems that it should be relatively straightforward to determine if direction is important in the representation of faces. For example, one prediction of such models is that there ought to be something 'special' about a given face's trajectory from the norm. That is to say, faces along a particular caricature trajectory should all be perceived as belonging to the same identity, with caricature exaggerating the identity information and anti-caricature attenuating the identity information. Most importantly, it is expected that lateral or oblique caricatures, which change the direction that a face deviates from the norm, will actually change the perceived identity. As a result, traditional norm-based accounts would seem to make a clear prediction. Namely, of the three types of caricature distortion it should be lateral or oblique caricatures that are the hardest to recognise as the original (undistorted) identity (see Section 1.3.3).

However, contrary to the predictions of the traditional norm-based model, it is actually anti-caricatures that are hardest to recognise followed by lateral or oblique caricatures with caricatures being recognised best (Lewis & Johnston, 1998; Rhodes et al., 1998). Indeed, these results are in line with an account based on exemplar density, whereby recognition of the target face is dependent on the ratio between the target exemplar's activation and the activation of other competing exemplars. Thus, while caricature would move a face into the lowest region of exemplar density, anti-caricature would move a face into the highest region of exemplar density, lateral or oblique caricatures would fall somewhere in between (e.g. Lewis & Johnston, 1999a).

Taken at face value, these results would seem to rule out a model in which direction of deviation from the norm has a role in the representation of identity. However, it is worth considering if there might be other reasons for the failure to find the anticipated pattern of results, particularly as some recent neurophysiological studies have leant weight to the direction-coding hypothesis (Leopold et al., 2006; Freiwald et al., 2009). Indeed, as other authors have pointed out (e.g. Lewis & Johnston, 1998; Stevenage, 1997), the lateral caricatures produced by Rhodes et al. were simple line-drawings and had a rather contorted appearance (see Figure 2.1), raising the possibility that there may have been something in the nature of their stimuli that might have inadvertently affected their findings.

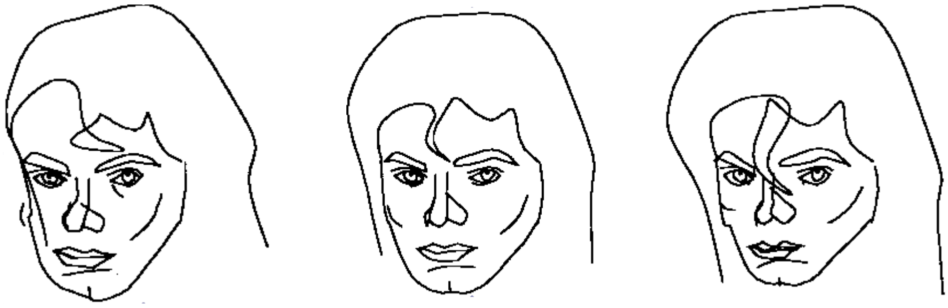


Figure 2.1. Morphed faces used by Rhodes et al. (1998). All morphs were created from line-drawings of famous people. Left image: lateral caricature, central image: anti-caricature, right image: caricature. Lateral caricatures appear to be more contorted than either anti-caricatures or caricatures.

To create the caricatures in their experiments, Rhodes et al. (1998) marked up corresponding points on both a target familiar face and an average face line-drawing. Each point on the target face was then moved by comparing its ‘x, y’ coordinates to the corresponding points on the average face and adding a percentage of the difference (see Brennan, 1985). When generating a caricature one can be relatively confident that, despite being a distortion of the veridical, the resultant face will not appear to be too unusual. This is because the caricature simply exaggerates the way in which a real face deviates from the average face. Thus, the morphing process is constrained by two points (average face and veridical) both of which are likely to lie within the psychological face-space.

However, when creating a Lateral caricature each of the points can be moved one of two ways (on the two-dimensional ‘x, y’ plane), that is to say, laterally to the right or left of the caricature trajectory for that point. The problem with this method of creating lateral caricatures is that there is no reason to suspect that the direction of distortion actually reflects the type of variance observed in a population of faces (Lewis & Johnston, 1998; Stevenage, 1997). That is to say, lateral caricatures are really only constrained by one reference point, with the lateral direction essentially an arbitrary distortion.

The effects of such unnatural distortions on face recognition are unclear. In one respect it would seem that applying an unnatural distortion to a face might cause it to appear more unusual and, thus, even less like the veridical image (Lewis and Johnson, 1998). However, this is not unambiguously the case. For example, if the distortion is in a direction that is not consistent with the variation normally seen in faces then it is essentially moving along a trajectory that is orthogonal to the manifold on which face identity is represented. As a result the actual change in the face space representation might actually be less than if the face was distorted in a more natural manner.

Lewis and Johnston (1998) went some way to addressing this issue. Using unfamiliar rather than familiar faces they investigated the effects of caricature, anti-caricature, and oblique caricature on judgments of similarity to the veridical image. Oblique caricatures are similar to lateral caricatures in that they do not lie on the trajectory between a given face and the average. The benefit of using oblique caricatures is that, rather than morphing points on a face in an arbitrary direction, it is possible to morph the points relative to a reference face. Specifically, Lewis and Johnston morphed the veridical images towards a reference face to create oblique caricatures, and towards and away from an average face to create caricatures and anti-caricatures. Lewis and Johnston found that caricatures were rated as most similar to the veridical identity, followed by the oblique caricatures and then the anti-caricatures. Again, as with the findings of Rhodes et al. these results indicate that the direction of deviation from the norm may not be psychologically important.

However, despite the absence of convincing behavioural evidence, the idea that facial identity might be captured by the direction of deviation from a norm is still prevalent in contemporary theoretical models of face processing (e.g. Giese and Leopold, 2005; Leopold & Bondar, 2005; Leopold et al., 2001; Loffler et al., 2005; Rhodes & Jeffery, 2006; Rhodes et al., 2005; Blanz, et al., 2000; Wilson, Loffler, & Wilkinson, 2002). Furthermore, while the focus of this thesis is on behavioural data, there has been some support for direction based coding from single cell recording studies of the macaque IT (Leopold et al., 2006), with the results seemingly fitting better with a traditional norm-based model than a standard exemplar-based model (Giese & Leopold, 2005). In addition, fMRI-adaptation in human participants has hinted that the fusiform face area (FFA) might show a greater release from adaptation, generally associated with greater dissimilarity, for faces on different trajectories from the norm than for faces lying on a single trajectory (Loffler et al., 2005).

In order to try and resolve the apparent discrepancy between neural recording and behavioural data, the studies presented in this chapter revisited the original caricature-versus-oblique caricature paradigm (e.g. Lewis & Johnston, 1998; Rhodes et al., 1998; Stevenage, 1997). As discussed, the traditional norm-based model predicts that faces along a given caricature trajectory will all activate the same population of face representations, whereas lateral or oblique caricatures ought to activate different populations of face representations. Thus, to directly test this hypothesis, the discriminability of face-pairs was measured both for face-pairs that lay on a caricature trajectory, traversing a midpoint, with each member of the pair lying on the same trajectory from the norm, and for face-pairs that lay on an oblique trajectory, traversing the same midpoint but lying on a trajectory that was oblique to the trajectory from the norm (see Figure 2.2).

2.2 Experiment 1: Discriminating Caricature and Oblique-Caricature in Familiar and Unfamiliar Upright Faces

Experiment 1, investigated discrimination thresholds for upright, familiar and unfamiliar Caricature - and Oblique-Trajectory face-pairs. The decision to use both familiar and unfamiliar faces was motivated by the fact that it is at least possible that they may differ in terms of how they are processed by the visual system (e.g. Burton & Jenkins, 2011). Indeed, some authors have even argued that unfamiliar faces are not processed as faces at all (Megreya & Burton, 2006). One possible interpretation of this is that unfamiliar face-pairs will be discriminated entirely on the basis of low-level image properties, whereas, familiar faces may be discriminated on the basis of higher-level differences. Indeed, this may be seen as a potential drawback of the study reported by Lewis and Johnston (1998), as they only used unfamiliar faces.

In addition, the specific method used to generate the stimuli in this experiment was chosen so as to create realistic stimuli that varied along trajectories that captured the natural variation found in faces. Thus, the oblique stimuli used in this study ought not to be subject to the same criticism that has been levelled at the study carried out by Rhodes et al. (1998). Moreover, the caricature and oblique trajectories were equated in terms of the physical differences they produce (a more detailed description is given in the methods) allowing for a more sensitive test than the method used by Lewis and Johnston (1998) as the physical difference between Oblique face-pairs and Caricature face-pairs was equated within subjects which it was not in their study[3].

To assess discriminability participants were shown three faces arranged on screen in a pyramid formation with one face, the target face, located centrally

---

[3] In Lewis and Johnston's (1998) study the oblique faces would, on average, have been slightly further from the veridical than the caricature trajectory faces which would have worked against their finding.

above the other two faces. One of the bottom two faces always matched the target and one face was slightly different from the target. Participants were then asked to identify which of the bottom two faces was the same as the top face. The bottom two faces were selected so that they traversed the midpoint relating to the 100% identity level of a given face. The maximum degree of separation between the two faces was 40% of the original identity (i.e. 20% either side of the midpoint) with the exact degree of separation used on a given trial chosen automatically using a Bayesian adaptive staircase procedure (see Tanner, Hill, Rasmussen, & Wichmann, 2005). Thus, unlike previous studies, which measured recognition of a familiar face (e.g. Rhodes et al., 1998) or the perceived similarity of two faces (Lewis & Johnston, 1998), this study provided a measure of the discriminability of face-pairs along either a caricature or an oblique trajectory

2.2.1 Methods

2.2.1.1 Participants: The participants were a sample of 44 undergraduates from Cardiff University School of Psychology who signed up for the experiment as partial fulfilment of course requirements. All participants reported that they had normal or corrected to normal vision.

2.2.1.2 Materials: Familiar and unfamiliar face stimuli were generated from 22 familiar and 22 unfamiliar full frontal, grey-scale photographs of male faces respectively. Unfamiliar faces were selected from a database held at Stirling University whereas the familiar face stimuli were selected from images found on the internet, selected on the basis of even lighting and high resolution. Familiar faces were all famous individuals chosen to be familiar to undergraduates.

Moreover, the assumption of familiarity was checked at the beginning of the experiment to ensure that they were familiar to all the participants. All images were cropped to remove any background and resized to a width of 350 pixels using Photoshop™. Corresponding points were then marked up on each of the faces so that they could be morphed to construct an average face using Psychomorph (Tiddeman, Burt, & Perrett, 2001). The faces were then separated into two sets (familiar and unfamiliar) of 11 arbitrarily chosen morph-pairs that were used to create the caricature- and oblique-trajectory pairs.

To create the Caricature-Trajectory pairs the faces were morphed both 20% towards and 20% away from an average face in steps of 2% using Psychomorph. Psychomorph essentially operates by dividing each face into a mesh of non-overlapping triangular tessellations the vertices of which are formed by three adjacent points (Benson & Perrett, 1991a). Caricaturing was carried out by first calculating the best fit between the points on the average face and the points on the face that was to be morphed. The difference between the 'x, y' location of each point on the average face and 'x, y' location of each of the corresponding points on the face to be morphed was then calculated and either exaggerated, to form a caricature, or attenuated, to form an anti-caricature.

The Oblique-Trajectory pairs were constructed separately for each of the morph pairs. For example, take the hypothetical morph pair Face 'A' and Face 'B' (see Figure 2.2). In this case the Oblique-Trajectory pairs would be constructed by taking the difference between the 'x, y' locations of the points on the average face and the 'x, y' locations on Face 'A'. Rather than apply this difference to Face 'A', as would be the case when constructing the Caricature-Trajectory pairs, this difference would be applied to Face 'B'. Thus, the points on Face 'B' would be moved in a direction and distance corresponding to the Caricature-Trajectory pair for Face ‘A’. However, this direction will be oblique to the Caricature-Trajectory of the Face ‘B’. To complete the counterbalancing the Caricature-Trajectory for Face ‘B’ would then be applied to Face ‘A’ so that the Oblique-Trajectory for Face ‘A’ was equivalent in distance and direction to the Caricature-Trajectory for Face ‘B’. Thus, across the full set of Caricature and Oblique-Trajectory pairs used in the experiment the stimuli were counterbalanced for morph distance, morph direction and the position of the morph midpoint. Furthermore, as the change applied to the faces to create the Oblique-Trajectory pairs represents the difference between a real face and the average face there is reason to believe that this should reflect the natural variation found in faces (see Figure 2.3 for examples of the stimuli).

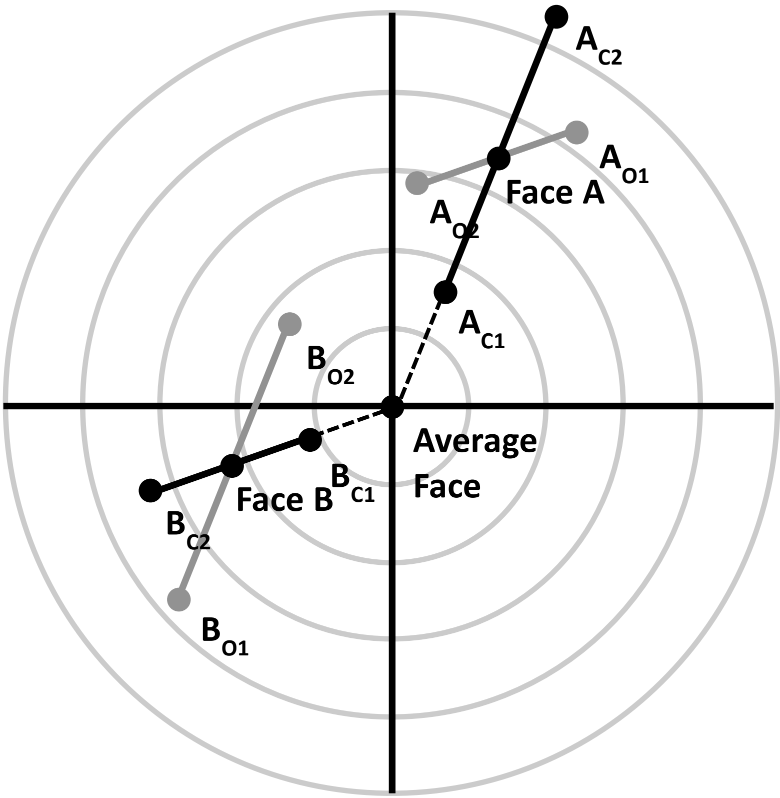


Figure 2.2. A simplified face-space diagram with Oblique-Trajectory pairs, A_{O1}-A_{O2} and B_{O1}-B_{O2} (represented in grey), and Caricature-Trajectory pairs, A_{C1}- A_{C2} and B_{C1}-B_{C2} (represented in black), traversing the hypothetical midpoint faces, Face A and Face B. Note that the trajectory A_{C1}- A_{C2} is the same length and is parallel to the trajectory B_{O1}-B_{O2}. Similarly, the trajectory B_{C1}-B_{C2} is the same length and is parallel to A_{O1}- A_{O2}. Thus, the average length and direction of the Oblique-Trajectory pairs is the same as the average length and direction of the two Caricature-Trajectory pairs.

Figure 2.3. Morphed unfamiliar Caricature- and Oblique-Trajectory face-pairs from Experiment 1. Top row: a Caricature-Trajectory pair separated by 40%. Bottom row: an Oblique-Trajectory pair separated by 40%.

2.2.1.3 Procedure: Participants were randomly assigned to one of two groups depending on whether they would see a familiar or an unfamiliar Caricature- and Oblique-Trajectory pairs. Within each of the two groups, each of the 22 participants was tested on a different one of the 22 test-pairs, whereby a test-pair consisted of one of the 11 Oblique-Trajectory pairs and the corresponding one of the 11 Caricature-Trajectory pairs (equivalent in distance). The faces were presented to participants on the screen in a pyramid formation as described previously. Participants were told that one of the bottom two faces would always match the top face and were asked to indicate which one it was. The faces were presented on a white background with the top image scaled to a width of 6cm and the two bottom images scaled to a width of 8cm with the bottom images slightly offset vertically. This was done in order to minimise the potential for a direct pixel-by-pixel comparison of any of the features of the face and to encourage an approach to the task that better reflected the mechanisms underlying face recognition.

On each trial the Bayesian staircase chose two faces, one on either side of the identity level being tested. Roughly speaking, the staircase algorithm works by updating its prior knowledge of the discrimination threshold (initially assuming all thresholds between 0% and 40% to be equally likely) based on the result of the previous trial. Following each trial the next step is then chosen so as to minimise the expected entropy of the posterior probability distribution, which is achieved by selecting the mean of the posterior distribution to test. These faces were presented as the bottom two images in the pyramid with the position of each face (i.e. left or right) randomly varied from trial to trial. The top face in the pyramid was chosen to match one the bottom two faces with the face that it matched randomly varied from trial to trial. Trials involving Caricature-Trajectory pairs and trials involving Oblique-Trajectory pairs were intermixed in a random order; however, the staircase program handled them on separate staircases. Participants indicated which of the bottom two faces they thought matched the top face by using the mouse to click on either a button marked “left” or a button marked “right”. If the participant did not respond within 5 seconds they were recorded as not knowing the answer and the next set of faces were displayed. The program stopped running after 100 trials had been completed for each of the trajectory types.

2.2.2 Experiment 1 Results

The discrimination thresholds, that is to say, the minimum degree of change as a percentage of identity required to reach 75% accuracy was recorded for each participant (mean thresholds for each condition are displayed in Figure 2.4). Thus, a lower discrimination threshold for a given condition the easier the face pairs were to discriminate. A repeated measures ANOVA with trajectory type (Caricature-Trajectory, Oblique-Trajectory) as the within subjects factor and familiarity (Unfamiliar, Familiar) as the between subjects factor revealed a significant main effect of trajectory type $F(1,42) = 12.248$, $p=.001$ with lower discrimination thresholds for the Oblique-trajectory pairs than for the Caricature-trajectory pairs. There was also a significant main effect of Familiarity $F(1,42) = 15.2$, $p<.001$ with lower discrimination thresholds for the familiar face pairs than for the unfamiliar face-pairs. There was no significant interaction between trajectory type and familiarity $F< 1$.

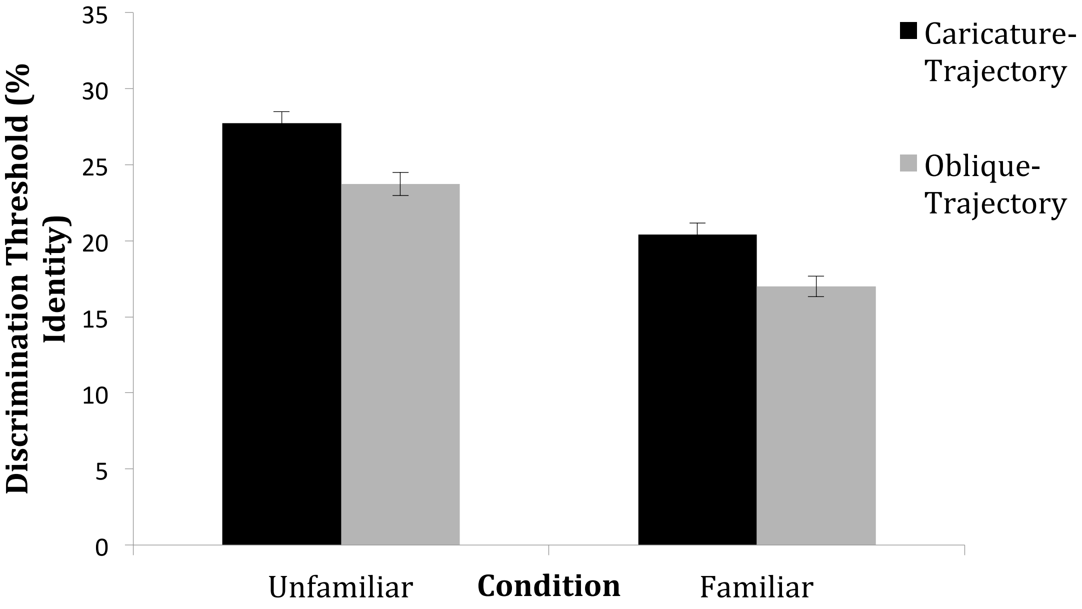


Figure 2.4. Discrimination thresholds for upright faces in Experiment 1 (% identity) vs. condition (Unfamiliar, Familiar) for Caricature- and Oblique-trajectory pairs (Error bars represent the SE in each condition).

2.2.3 Experiment 1 Discussion

Experiment 1 aimed to investigate the specific role of the norm in the representation of faces. Specifically, Experiment 1 investigated whether direction relative to the norm is important in the representation of faces, in line with traditional accounts of the norm-based model. Participants' discrimination thresholds were measured for face-pairs separated along either a Caricature- or Oblique-Trajectory. Given that Oblique-Trajectory face-pairs have different directions of deviation from the norm, traditional norm-based models would suggest that they should be more easily discriminated than Caricature-Trajectory pairs. In line with such accounts, the results indicated that participants' discrimination thresholds were significantly lower for Oblique-Trajectory pairs than for Caricature-Trajectory pairs.

These results are in contrast to the previously reported behavioural results (e.g. Rhodes et al., 1998; Lewis & Johnston, 1998). However, they do appear to be consistent with the currently emerging picture from fMRI-adaptation studies of the FFA (Loffler, et al., 2005) and single cell recording studies in monkeys (Leopold et al., 2006). It is difficult to pinpoint exactly what may have led to the different results obtained in this study as a number of factors differ between this study and the other studies investigating lateral and oblique caricature. Clearly, the stimuli used in this study are more naturalistic than those used by Rhodes and colleague. However, given that Lewis and Johnston (1998) used fairly naturalistic face stimuli, created using a similar rationale to the ones used here, it is unlikely that this would be the only cause of the discrepant results.

It was suggested in the introduction that one reason that Lewis and Johnston may not have found an effect of direction in their study is that they used unfamiliar faces. Indeed, the fact that some authors have suggested that unfamiliar faces are not processed as faces (Megreya & Burton, 2006) was the rationale behind including both familiar and unfamiliar faces in the present study. However, while the results reported here reveal a significant main effect of familiarity on discrimination thresholds there was no interaction of familiarity and trajectory type (Caricature or Oblique). Moreover, in light of the failure to find an interaction it would seem that one should be cautious about interpreting the lower thresholds found for familiar faces as they were taken from a very different stimulus set. It may be, for example, that the familiar faces had more uneven lighting than the unfamiliar faces making them appear more distinctive (though every effort was taken to avoid this).

Another possibility is that the nature of the task used in Experiment 1, discrimination verses likeness rating or recognition, led to the difference between Caricature- and Oblique-Trajectory pairs observed in this experiment. An atypicality bias has been observed in previous experiments for both faces and other objects, suggesting that participants tend to rate visual stimuli as more similar to distinctive rather than typical category exemplars (e.g. Tanaka & Corneille, 2007). With this in mind it seems possible that measuring discriminability might provide a less biased measure of the underlying neural representation of faces, perhaps explaining why the results obtained here are more in line with the neural data. Nevertheless, given the discrepancy between the results obtained here and the results reported previously, it seems prudent to rule out the possibility that unintended low-level visual differences between the Caricature- and Oblique-Trajectory pairs might have contributed to these findings.

2.3 Experiment 2: Discriminating Caricature and Oblique Caricature in Familiar and Unfamiliar Inverted Faces

Experiment 2, investigated discrimination thresholds for inverted, familiar and unfamiliar Caricature - and Oblique-Trajectory face-pairs. There is a great deal of work in the face recognition literature suggesting that inverted faces are recognised less holistically and more featurally than upright faces (e.g. Young, Hellawell & Hay, 1987). Furthermore, Rhodes (1996) has argued that the recognition advantage attributed to caricatures may only be apparent for upright faces and other objects of expertise (e.g. Rhodes & McLean, 1990; Rhodes & Tremewan, 1994). The fact that at least some of the mechanisms of face expertise do not transfer to inverted faces might lead one to speculate that inverted faces would not be encoded in terms of their direction of deviation from the norm. However, on this point it is worth noting that many of the results implicating a norm-based model of face recognition have been successfully replicated using inverted faces (Leopold et al., 2001; Robbins et al., 2007). Thus, although it is difficult to make any clear prediction regarding the discriminability of inverted Oblique- and Caricature-trajectory pairs the results ought to inform us about the relationship between norm-based coding and other measures of expertise.

2.3.1 Methods

2.3.1.1 Participants: The participants were a sample of 44 undergraduates from Cardiff University School of Psychology who signed up for the experiment as partial fulfilment of course requirements. All participants reported that they had normal or corrected to normal vision.

2.3.1.1Materials & Procedure: Both the materials and procedure used in Experiment 2 were identical to those in Experiment 1 except that the faces were inverted, 180° rotation, using Adobe Photoshop™.

2.3.2 Experiment 2 Results

The discrimination thresholds were recorded for each participant as in Experiment 1, with a lower discrimination threshold for a given condition indicating that the face-pairs were easier to discriminate (mean thresholds for each condition are displayed in Figure 2.5). A repeated measures ANOVA with trajectory type (Caricature-Trajectory, Oblique-Trajectory) as the within subjects factor and familiarity (Unfamiliar, Familiar) as the between subjects factor revealed there was no significant main effect of trajectory type $F(1,42)=1.315$, $p=.558$, with Oblique-Trajectory pairs no easier to discriminate than Caricature-Trajectory pairs. There was a marginally significant main effect of Familiarity $F(1,42)= 3.729$, $p=.06$ with lower discrimination thresholds for the familiar face pairs than for the unfamiliar face-pairs. The interaction between trajectory type and familiarity was not significant $F<1$.

2.3.3 Combined Experiment 1 & 2 Results

A combined analysis was also conducted on the data from Experiments 1, 2 with a repeated measures ANOVA with direction (Caricature-Trajectory, Oblique-Trajectory) as the within subjects factor and orientation (Upright, Inverted) and familiarity (unfamiliar, familiar) as the between subjects factors. This analysis revealed a significant main effect of direction $F(1,84)=8.625$, $p=.004$ with lower discrimination thresholds for the Oblique-Trajectory pairs than for the Caricature-Trajectory pairs as well as a significant between subjects main effect of familiarity $F(1,84)=15.408$, $p<.001$ and orientation $F(1,84)=11.305$, $p=.001$. There were no significant interactions between either direction and orientation $F(1,84)=1.107$, $p=.296$, or orientation and familiarity $F<1$ and the three way interaction between direction, familiarity and orientation was also not significant $F<1$.

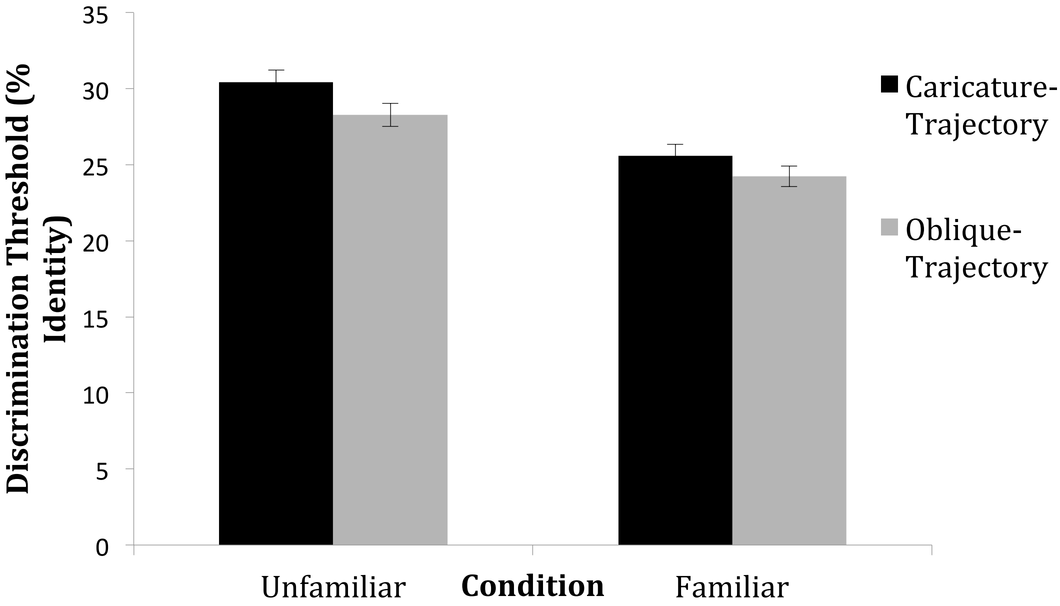


Figure 2.5. Discrimination thresholds for inverted faces in Experiment 2 (% identity) vs. condition (Unfamiliar, Familiar) for Caricature- and Oblique-trajectory pairs (Error bars represent the SE in each condition).

2.3.4 Experiment 2 Discussion

Experiment 2 aimed to expand on the results of Experiment 1, investigating whether the direction encoded relative to the norm is important in the representation of inverted faces. Thus, as in Experiment 1, lower discrimination thresholds for Oblique-Trajectory face-pairs than for Caricature-Trajectory face-pairs would suggest that direction was important in the encoding of inverted faces. However, the results revealed that there was no effect of trajectory direction on the discrimination thresholds for inverted faces. There was, however, a marginally significant between-subjects main effect of familiarity, as there was in Experiment 1. Although, as stated previously this effect ought to be interpreted cautiously given that the familiar face images used in this experiment were not taken under the same tightly controlled conditions as the unfamiliar face images.

Despite the different pattern of statistical significance seen when each of Experiments 1 and 2 are analysed separately, a combined analysis of the Experiment 1 and 2 results reveals a main effect of trajectory direction but no interaction with orientation or familiarity. Thus, although the general difference between caricature and oblique trajectories is clear, the results of Experiment 2 are somewhat inconclusive with regards to whether inverted faces are encoded with respect to their direction of deviation from the norm. This is perhaps not particularly surprising, as it is possible that the representation of the inverted faces is simply more impoverished than the representation of upright faces. Supporting this hypothesis, the significant main effect of orientation suggested that discrimination was more difficult for inverted than upright faces. Indeed, although it might be argued that some other face specific, or expertise specific, phenomena such as holistic processing are not seen for inverted faces (e.g. Robbins & McKone, 2003; though see Richler, Mack, Palmeri, & Gauthier, 2011) it appears that many of the reported norm-based adaptation aftereffects are present in both upright and inverted faces (Leopold et al., 2001; Robbins et al., 2007).

2.4 General Discussion

The two studies reported here investigated whether faces are encoded by way of their direction of deviation from the norm, as is suggested by traditional versions of the norm-based model. Consistent with this idea, a combined analysis of the results from Experiments 1 and 2 revealed that discrimination thresholds were significantly lower for Oblique-trajectory pairs (which lie on different trajectories from the norm) than Caricature-trajectory pairs (which lie on the same trajectory from the norm).

It is interesting to consider in more detail why it is that previous studies did not find the same pattern of results as was found here (e.g. Lewis & Johnston, 1998; Rhodes et al., 1998). As discussed earlier, similarity judgements, such as were used in Lewis and Johnston's task, appear to be strongly biased by exemplar distinctiveness (Tanaka & Corneille, 2007). It is therefore possible that the task used in the present study reduced such bias, as participants were engaged in making a perceptual match with all images presented on screen simultaneously. In addition, in Lewis and Johnston's study the target face was presented a few seconds before the two foil faces which may have increased any bias by increasing memory load. Indeed, this might be seen as a drawback of the present study as some previous research has suggested that perceptual matching tasks such as the Benton Facial Recognition Test (Benton, Sivan, Hamsher, Vamey, & Spreen, 1983) do not provide a good measure of face specific recognition ability (e.g. Duchaine & Weidenfeld, 2003).

In any case, even if perceptual matching did occur it would be difficult to see how this could have resulted in the difference seen here between Oblique- and Caricature-Trajectory face-pairs since the trajectories were matched on the physical distance of the morphs. Indeed, the stimuli were counterbalanced for the position of the midpoint face in face-space, the average physical distance between the face-pairs, and the direction of each trajectory in face-space, making low-level explanations unlikely.

While low-level artefacts are unlikely to have influenced the current results it is possible that there may be high-level differences between the two trajectory types (not relating to direction of deviation from the norm). For example, the Caricature- and Oblique-Trajectory pairs may differ in the degree to which they affect the most salient or distinctive features of the faces. For example, if one were to take a face that was fairly average except for it having a large nose then most of the variation between caricature morphs would be explained by changes in the size of the nose. Thus, a participant using a featural strategy could pay attention to some particularly distinctive feature in order to discriminate the morph pairs. In contrast the oblique morph trajectory may produce very little variation in the size of the nose, instead affecting a different feature or set of features. However, this issue would actually seem to predict the opposite pattern of results to those observed. That is to say, it predicts that

Caricature-Trajectory pairs would be more discriminable than Oblique-Trajectory pairs.

In conclusion, while studies into face adaptation have provided support for the idea that faces are encoded relative to a norm (e.g. Leopold et al., 2001; Rhodes & Jeffery, 2006; Robbins et al., 2007) it is still unclear what role the norm plays in the representation of faces. The two experiments reported in this chapter provide the first behavioural support for the claim that direction relative to the norm is psychologically important for the representation of faces. Unfortunately, as a result of the numerous differences between the current experiments and previous studies it is somewhat difficult to determine exactly which factors may have contributed to the discrepancy between these results and the results of previous studies (e.g. Lewis & Johnston, 1998). It is worth noting here that, in light of the results reported in subsequent chapters, an alternative, and perhaps more parsimonious, account of these results will be considered in Chapter 6.

Chapter 3

Discrimination Across the Norm

Abstract

Chapter 3 reports two experiments that investigate whether discrimination of face pairs that lie on different sides of the average face differs from discrimination at other points along the caricature trajectory. In Experiment 3 the discrimination thresholds for unfamiliar face-pairs at various distances along trajectories up to and across the average face were measured. The results indicated that there was no effect of the distance of a face-pair from the average on discrimination threshold. Importantly, contrary to the predictions of the traditional norm-based model, in which direction of deviation from the norm is encoded, face-pairs that traversed the average face were not more easily discriminated than face-pairs at any other point along the caricature trajectory. In Experiment 4 discrimination thresholds for familiar face-pairs at various distances along trajectories up to and across the average face were measured. As in Experiment 3 there was no effect of distance from the average on discrimination threshold. Moreover a combined analysis of Experiments 3 and 4 did not reveal any significant effect of distance from the average on discrimination threshold. The results are considered in terms of different accounts of face-space, suggesting that direction relative to the norm is not directly relevant for the representation of faces in the visual system.

3.1 Introduction

As discussed in Chapter 2, recent accounts of the norm-based model have speculated that the visual system might represent facial identity by the direction that a given face deviates from the norm (e.g. Giese and Leopold, 2005; Leopold et al., 2006; Leopold et al., 2001; Loffler et al., 2005; Rhodes et al., 2005; Blanz et al., 2000; Wilson et al., 2002). Chapter 2 provided some behavioural evidence to support this position, demonstrating that discrimination thresholds were lower for oblique-trajectory face-pairs than for face-pairs that lay on the caricature-trajectory. This result indicates that all directions in face-space are not equally discriminable, implicating a role of the angle, referenced against the norm, in the discriminability of a pair of faces.

If the angle, or difference in direction from the norm, between two faces is indeed relevant to their discriminability then the effect ought to be strongest for face-pairs that lie on different sides of the norm, that is to say, faces that are separated by an angle of 180^{o}. Some authors have speculated that faces on different sides of the norm will actually be perceived as opposites (Blanz et al., 2000). A number of previous studies have investigated discrimination thresholds along a trajectory up-to and across the norm, with very mixed results (e.g. Blanz et al., 2000; Dakin & Omigie, 2009; Rhodes et al., 2007; Wilson et al., 2002). Indeed, a recent review of the findings across such studies concluded that there was evidence of multiple different encoding strategies but could not identify any clear indication as to what experimental factors might lead to the different results obtained across studies (Susilo et al., 2010b).

For example, Rhodes et al. (2007, Experiment 1) measured the ability of participants to discriminate changes in interocular distance in faces that lay at different distances up-to and across the average face. The faces in this study were morphed along their respective caricature trajectories between the veridical image and an average in steps of 10% of the original identity. At each level of identity, the faces were manipulated, varying the interocular distance to create stimuli that lay on a trajectory oblique to the caricature trajectory. If, as indicated by the findings reported in Chapter 2, faces are encoded by their distance and direction of deviation from the norm then, given that the angle between the norm and two oblique face-pairs will be greatest for oblique face-pairs close to the norm, the discrimination threshold ought to be lowest close to the average face. However, the results revealed no peak in the discrimination threshold either around the norm or at any other point along the caricature trajectory. While these results would seem to run contrary to the findings reported in Chapter 2, a similar study, reported by Wilson et al. (2002) indicated that, in line with a traditional norm-based model, the discrimination threshold for oblique trajectory face-pairs was in fact lower for faces that were closest to the norm.

While the results of these two studies are rather contradictory, neither study actually measured discrimination thresholds for face-pairs that lay on the same caricature trajectory. As a consequence, the distinctiveness of the faces and the angle between the face-pairs being discriminated may be confounded. That is to say, faces that are separated by a very large angle, such as when the interocular distance manipulation is applied to a typical face, are necessarily very typical. In contrast, faces that are separated by a very large angle, such as when the interocular distance manipulation is applied to a distinctive face, are necessarily very distinctive. Thus, if as predicted by a traditional norm-based model, direction relative to the norm is important one might expect that discrimination thresholds would decrease in proportion to how average a given face is. But, it is also possible that more average faces are more difficult to discriminate as they produce weaker activation in norm-coded face representations (i.e. there would be a greater proportion of noise in the signal). These two effects may cancel out in some cases causing the appearance of no change in the discrimination threshold at different points up to and across the norm.

As it turns out, distinctiveness alone does appear to influence the discriminability. For example, Dakin and Omigie (2009) measured the discriminability of face-pairs that lay at various points along a caricature trajectory between an average face and a 125% caricature. They reported that discrimination along the caricature trajectory follows a roughly 'U' shaped function whereby discrimination close to the norm is more difficult than discrimination at an intermediate level of caricature. Unfortunately, Dakin and Omigie did not measure the discriminability of face-pairs that actually traverse the average face. Indeed, for the present line of research this point is the most interesting, as it is the only point along the caricature trajectory for which the face-pairs lie on different trajectories from the norm. Thus, if the angle between each member of a face-pair and the norm is indeed relevant to the psychological representation of similarity then there ought to be a distinct change in the discrimination threshold function for pairs of faces traversing the norm.

Thus far, two studies have measured the perceived similarity, but not discriminability, of face-pairs lying on a caricature trajectory extending from a given face, up-to and across the average face with quite contradictory results (Blanz et al., 2000; Rhodes et al., 2007, Experiment 2, 3, & 4). Rhodes et al. obtained ratings of perceived similarity for face-pairs that lay on the same caricature trajectory separated by either 40% or 20% of the full identity morph. Their results indicated that similarity was greatest at the norm, a result that is inconsistent with the hypothesis that direction is important. Contrastingly, and in line with traditional norm-based accounts, Blanz et al. (2000) have suggested that perceived similarity is actually reduced for face pairs spanning the norm. Unfortunately, Blanz et al. results are of questionable reliability given that the step size in caricature across the average face was twice as large as the step size used elsewhere in the space. Indeed, Blanz et al. assumed that the ratings for similarity, given on a three-point scale, should be twice as high for the step across the average, reflecting the fact that the step size was twice as large - an assumption that would only hold if similarity ratings were linearly related to step size.

The goal of the two experiments presented in this chapter was to extend on the results of Dakin and Omigie (2009), measuring discrimination thresholds for faces at various points along the caricature trajectory including thresholds for face-pairs that traverse the average face. While Dakin and Omigie reported that discrimination followed a roughly ‘U’ shaped function with discrimination thresholds largest around the average, they did not test discriminability for face-pairs traversing the average face. Indeed, at present the results of previous studies that have measured similarity for face-pairs across the norm are somewhat inconsistent. That is to say, while Rhodes et al. (2007) reported that face-pairs that traverse the average are perceived as more similar and Blanz et al. (2000) reporting that they are perceived as less similar.

3.2 Experiment 3: Discriminating Unfamiliar Faces Across the Norm

Experiment 3 investigated discrimination thresholds for unfamiliar face-pairs at various positions along the caricature trajectory up-to and across the average face. Unlike in previous experiments (e.g. Rhodes et al., 2007, Experiment 1; Wilson et al., 2002), discrimination thresholds were measured for face pairs that both lay on the same caricature trajectory (i.e. they were not on an oblique trajectory) removing the potentially confounding effect of the angle relative to the norm that was present in previous studies. In addition, in light of the results reported in Chapter 2, it was thought that measuring discrimination thresholds, rather than perceived similarity (e.g. Rhodes et al., 2007, Experiments 2, 3, & 4) might provide a better measure of the underlying face-space representations.

The method used to measure discriminability was the same as that used in the two studies reported in Chapter 2. That is to say, faces were arranged on screen in a pyramid formation with one face, the target, located centrally above the other two faces. One of the bottom two faces always matched the target and one face was slightly different from the target. The bottom two faces were selected so that they traversed the midpoint of the identity level being tested (0%, 40%, 80% & 120%) with the maximum degree of separation between the two faces set to 40% of the original identity. Therefore, at the 0% identity level, face pairs were selected from opposite sides of the average (0% identity). If direction is important in the psychological representation of faces then discrimination thresholds at 0% identity ought to be lower than at other points along the continuum. A final possibility to consider is that if discrimination thresholds increase with distinctiveness it might be the case that discrimination across the norm is not the lowest along the caricature trajectory. If this is the case then it should be apparent in the shape of the discrimination function and may require further testing to see if there is a discontinuity in the function across the average face.

3.2.1 Method

3.2.1.1 Participants: The participants were a sample of 22 undergraduates from Cardiff University School of Psychology who signed up for the experiment as partial fulfilment of course requirements. All participants reported that they had normal or corrected to normal vision.

3.2.1.2 Materials: The face stimuli were 22 full frontal, grey-scale photographs of unfamiliar male faces selected from a database held at Stirling University. All images were cropped to remove background and resized to a width of 350 pixels using Photoshop™. Corresponding points were then marked up on each of the faces so that they could be morphed to construct an average face using Psychomorph (Tiddeman et al., 2001). Four morph parents were then constructed for each of the 22 faces by morphing them towards or away from the average face to create a 0%, 40%, 80% and 120% identity level. Finally, each face at each identity was morphed by 20% of the original (100%) identity towards and away from the average in 2% steps (maximum separation of 40%).

3.2.1.3 Procedure: Each participant was tested on a different face for each of the four identity levels. This was done in order to minimize the chance of a participant learning a specific feature of the face that might make discrimination particularly easy. The faces were selected so that over the 22 participants each of the 22 faces was tested exactly once at each identity level. The faces were presented to participants on the screen in a pyramid formation as described previously (see Experiment 1 for details of the presentation).

On each trial the staircase program (see Experiment 1) chose two faces, one on either side of the identity level being tested. These faces were presented as the bottom two images in the pyramid with the position of each face (i.e. left or right) randomly varied from trial to trial. The top face in the pyramid was chosen to match one of the bottom two faces, with the face that it matched randomly varied from trial to trial. Trials at each of the four identity levels/faces were intermixed in a random order; however, the staircase program handled them on separate staircases. Participants indicated which of the bottom two faces they thought matched the top face by using the mouse to click on either a button marked “left” or a button marked “right”. If the participant did not respond within 5 seconds they were recorded as not knowing the answer and the next set of faces were displayed. The program stopped running after 100 trials had been completed for each of the four identity levels/faces.

3.2.2 Experiment 3 Results

The mean discrimination threshold at each midpoint identity level was calculated across participants and plotted as a function of identity (Figure 3.1). An examination of the plot suggested that there was no peak in the discrimination function around the average or at any identity level. This was confirmed by a one-way repeated-measures ANOVA on discrimination threshold with level of identity level (0%, 40%, 80% and 120%) as the repeated measures factor, $F<1$.

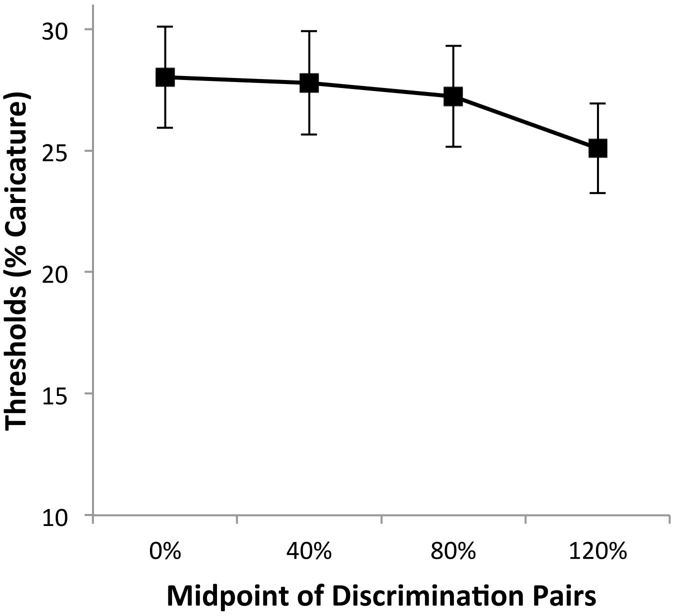


Figure 3.1: Mean discrimination threshold averaged across participants vs. Midpoint identity level for unfamiliar faces in Experiment 3 (0% represents discrimination for pairs spanning the average face). Error bars represent the SE in each condition.

3.2.3 Experiment 3 Discussion

In Experiment 3 the discrimination threshold of unfamiliar faces at various positions along the caricature trajectory was assessed. If direction of deviation from the norm is important for the psychological representation of faces then discrimination thresholds should be lower for face-pairs that traverse the average than for face-pairs traversing any other midpoint along the caricature trajectory. Indeed, face-pairs that traverse the norm will actually be separated by the greatest possible angle relative to the norm, meaning that they should in some sense be perceived as opposites (Blanz et al., 2000). However contrary to these predictions, the results of Experiment 3 indicated that there was no effect of identity level on discrimination threshold. There was also no evidence for the alternative possibility, considered in the introduction, that discriminability might increase for face-pairs close to the norm.

These results are rather surprising given that Chapter 2 reported two studies that appeared to demonstrate that direction is important for our psychological representation of faces. However, it should be noted that the present results are actually more inline with other research in the literature. Indeed, Rhodes et al. (2007, Study 1) reported that discrimination thresholds were not affected by identity level, and other studies (e.g. Dakin & Omigie, 2010; Rhodes et al; Studies 2, 3, & 4) have actually reported that, if anything, faces that traverse the norm are perceived as more similar. Moreover, research into the effects of lateral and oblique caricature on face recognition has (apart from the studies reported in this thesis) not found any evidence that direction is important (Lewis & Johnston, 1998; Rhodes et al., 1998).

One factor that is worth considering before drawing any strong conclusions is that the faces used in Experiment 3 were all unfamiliar to the participants. Indeed, the experiment was set up so as to reduce the amount of familiarisation that would occur throughout the study by testing the participants with a different face for each identity level. As was discussed in the introduction to Experiment 1 (Chapter 2) there has been some suggestion in the literature that unfamiliar faces may differ in terms of how they are processed by the visual system (e.g. Burton & Jenkins, 2011), with some authors having argued that unfamiliar faces are not processed as faces at all (Megreya & Burton, 2006). Thus, while there did not appear to be any evidence that familiar and unfamiliar faces were processed differently in either Experiment 1 or 2, it is worth ruling out the possibility that the findings might be different if familiar faces were used.

3.3 Experiment 4: Discriminating Familiar Faces Across the Norm

Experiment 4 sought to further investigate whether the direction of deviation from the norm is important for the psychological representation of faces. As was discussed earlier in the thesis there is some suggestion that unfamiliar faces may differ in terms of how they are processed by the visual system. Thus, Experiment 4 sought to replicate the results of Experiment 3 using familiar faces. The procedure used in Experiment 4 was identical to that used in Experiment 3 except that familiar celebrity faces were used rather than unfamiliar faces. In addition, participants were shown the four faces that they would be tested on before beginning the experiment to ensure that they recognised them and that they were fully familiarised with them.

3.3.1 Method

3.3.1.1 Participants: The participants were a sample of 22 undergraduates from Cardiff University, School of Psychology who signed up for the experiment as partial fulfilment of course requirements. All participants reported that they had normal or corrected to normal vision.

3.3.1.2 *Materials:* The experimental stimuli were constructed in the same way as for Experiment 3. They were constructed from 22 photographs of well-known male celebrities selected from images found on the internet. The stimuli were selected so that they were all full frontal photographs with relatively even lighting and a high resolution. In addition the participants' familiarity with the faces was checked prior to the start of the study.

3.3.2 Experiment 4 Results

The mean discrimination threshold at each midpoint identity level was calculated across participants and plotted as a function of identity (Figure 3.2). An examination of the plot suggested that there was no peak in the discrimination function around the average or at any identity level. This was confirmed by one-way repeated-measures ANOVA on discrimination threshold with level of identity level (0%, 40%, 80% and 120%) as the repeated measures factor, $F<1$.

3.3.2.1 Combined Experiment 3 & 4 Results

A combined analysis was also conducted on the data from Experiments 3 and 4 with a repeated measures ANOVA with identity strength as the within subjects factor and familiarity (unfamiliar, celebrity) as the between subjects factor. The ANOVA revealed a significant main effect of familiarity (unfamiliar or celebrity) $F(1, 42) = 4.07$, $p = .05$ with lower discrimination thresholds for celebrity faces (mean= 22.56% caricature) than for unfamiliar faces (mean= 27.52% caricature), but no main effect of identity level, $F<1$. Furthermore, there was no interaction, $F<1$.

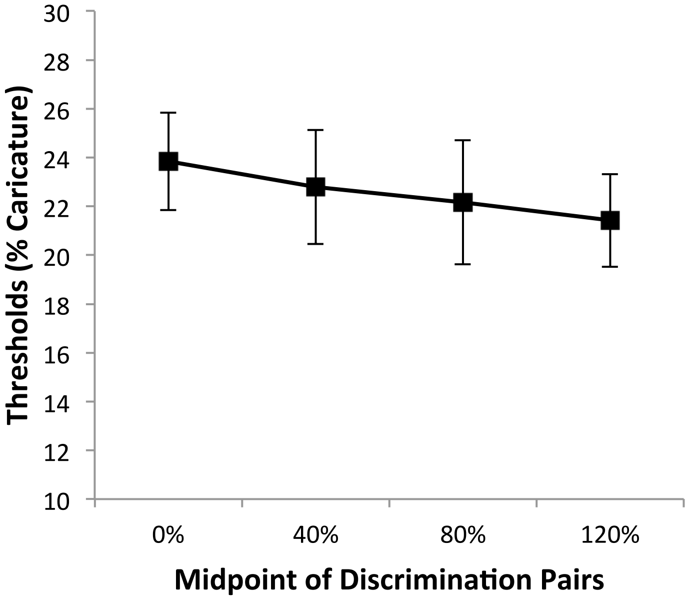


Figure 3.2: Mean discrimination threshold averaged across participants vs. midpoint identity level for familiar faces in Experiment 4 (0% represents discrimination for pairs spanning the average face). Error bars represent the SE in each condition.

3.4 Experiment 4 Discussion

Experiment 4 aimed to expand on the results of Experiment 3, measuring discrimination thresholds for familiar (celebrity) faces at various positions along the caricature trajectory. If direction of deviation from the norm were important for the psychological representation of faces, then it would be expected that discrimination thresholds should be lower for face-pairs that traverse the average than for face-pairs at other points along the caricature trajectory. However contrary to this prediction, but in line with the results of Experiment 3 the results of Experiment 4 indicated that there was no effect of identity level on discriminability. Thus, these findings do not support the hypothesis that direction relative to the norm is important in the representation of identity.

In addition a combined analysis of the data from Experiments 3 and 4 revealed that while there was a significant main effect of familiarity (unfamiliarity vs. celebrity) on discrimination thresholds, with lower discrimination thresholds for the celebrity faces compared to the unfamiliar faces, there was no overall main effect of identity level. These results replicate the finding from Experiments 1 and 2 that the familiar face images were more easily discriminated than the unfamiliar face images. However, as was mentioned previously it is difficult to read too much into these results as the images were taken from very different sources.

In addition, as there did not seem to be an effect of caricature level on discriminability either in the analysis of Experiment 4 or in the combined analysis it seems unlikely that decreasing discrimination thresholds close to the average were simply obscuring the effect of crossing the average from being detected. Indeed, surely if faces on opposite sides of the average face really are perceived as psychological opposites then the effect would not be so subtle.

3.5 General Discussion

The two experiments reported in Chapter 3 did not find any evidence to support the claim that faces are encoded by way of their direction of deviation from the norm. Specifically, discrimination thresholds were measured for unfamiliar (Experiment 3) and familiar (Experiment 4) faces at various points along a caricature trajectory up-to and across the norm. If direction were important to the representation of faces, as was suggested by the studies reported in Chapter 2, it would be expected that the discrimination thresholds would be lower for face-pairs that traversed the norm than for face-pairs traversing a midpoint at any other point along the caricature trajectory. However, neither the individual analyses of Experiment 3 and 4 or the combined analysis revealed such a pattern of results.

While these results are not in line with the interpretation of the results in Chapter 2 they are consistent with previous findings in the literature (e.g. Rhodes et al., 2007, Study 1). Moreover, in as far as these results indicate that direction of deviation from the norm is not important in the representation of faces they are also broadly consistent with findings from the lateral/oblique caricature literature (Lewis & Johnston, 1998; Rhodes et al., 1998). Indeed, with the exception of the single, possibly flawed (see Section 3.1), study reported by Blanz et al. (2000) there would not appear to be any evidence for the claim that faces which traverse the norm are perceived as ‘opposites’. On the contrary, some previous research has actually suggested that face-pairs traversing the average might be perceived as more similar (e.g. Dakin and Omigie, 2009; Rhodes et al., 2007, Studies 2, 3, & 4).

It is worth briefly considering the results of Chapters 2 and 3 in the context of the different versions of the face-space model discussed in the introductory chapter. As has been discussed, while there is currently substantial evidence that faces are encoded with respect to a norm (see Rhodes & Leopold, 2011, for a review), the precise way in which such norm coding is realised is still relatively unclear. To this end, the studies in Chapters 2 and 3 set out to investigate whether direction relative to the norm was important in the representation of faces, aiming to differentiate traditional accounts of the norm-based model from alternative versions. However, in this respect the studies in this Chapter appear to contradict the findings in Chapter 2.

So, is direction relative to the norm important? One possibility is that direction in face-space is not directly encoded by individual face representations, as it would be in a traditional norm-based model. Rather, direction might be realised by the opponent encoding of pairs of overlapping face representations, as is suggested by the two-pool model. If this were the case then while the transition across the norm may not signal a sudden change in perceived identity (it would be a gradual transition from one pool to another), it would still be possible that direction is important. Assuming that face representations are more broadly-tuned in the direction across the norm than in lateral or oblique directions it might be possible to account for the findings in both Chapter 2 and Chapter 3.

In summary, the two experiments reported in Chapter 3 did not find that discrimination for face-pairs that lie on different sides of the average face differs from discrimination at other points along the caricature trajectory. While it seems that the results of Chapter 3 would be inconsistent with a model in which individual face representations respond preferentially to particular directions of deviation from the norm (e.g. Giese & Leopold, 2005), they are still consistent with alternative accounts of the norm-based model, such as the two-pool norm-based model. However, it should be noted that, taken alone, the results of Chapter 3 would also be consistent with an exemplar-based account. That is to say, discriminability would simply be related to the Euclidean distance between two faces. Note that an interpretation such as this will be discussed in more detail in Chapter 6.

Chapter 4

Modeling Face Adaptation

Abstract

Chapter 4 reports a series of studies designed to investigate the effects of adaptation in norm- and exemplar-based models. Face adaptation was implemented in three computational instantiations of the face-space model, a traditional norm-based model (e.g. Giese & Leopold, 2005), a two-pool norm-based model (e.g. Rhodes & Jeffery, 2006), and an exemplar-based model. In Experiment 5a a version of the paradigm used by Leopold et al. (2001) was instantiated; the results revealed that all three models made predictions that were qualitatively consistent with the findings reported in the literature. Moreover, Experiment 5b and 5c, revealed that aftereffects in all three models appeared to be centred on the average of the population of faces, a finding that is normally considered an exclusive prediction of the norm-based model. In Experiment 6 a version of the task used by Rhodes and Jeffery (2006) was implemented; the results revealed that both the two-pool norm-based model and the exemplar-based model made predictions that were qualitatively consistent with the findings reported in the literature. However, the traditional norm-based model failed to make qualitatively accurate predictions over a large range of parameters. Finally, in Experiment 7, a version of the task reported by Robbins et al. (2007) was implemented. Again, all three models made qualitatively accurate predictions across a wide range of parameters. Together the results of these five experiments suggest that, contrary to previous claims in the literature, the evidence in favour of a two-pool model, and against the exemplar-based model, is far from conclusive.

4.1 Introduction

Recent research into face adaptation is widely accepted as providing support for a norm-based model and against an exemplar-based model (e.g. Freiwald et al., 2009; Griffin, et al., 2011; Leopold & Bondar, 2005; Leopold et al., 2001; Rhodes & Leopold, 2011; Rhodes & Jaquet, 2011; Rhodes, et al., 2011; Rhodes & Jeffery, 2006; Rhodes et al., 2010; Rhodes et al., 2005; Robbins, et al., 2007; Susilo, et al. 2010b; Tsao & Freiwald, 2006), with some authors even suggesting that the evidence is now 'conclusive' (Susilo, et al., 2010a, p. 300). Furthermore, the use of established adaptation paradigms to implicate norm-based coding in other areas of face recognition is widespread. Recent studies have implicated norm-based coding in emotion perception (Skinner & Benton, 2010) as well as providing evidence that even young children have well established face norms (Anzures et al., 2009; Jeffery et al., 2010; Nishimura et al., 2008; Nishimura et al., 2011; Pimperton et al., 2009; Short et al., 2011). In addition, face adaptation has been used to gain a better understanding of developmental disorders such as autism (Pellicano et al., 2007) and congenital prosopagnosia (Nishimura et al., 2010; Palermo et al., 2011) using the norm-based model as a basis for theorising about the possible underlying causes of such conditions.

As was discussed previously (Section 1.4.1), the widespread acceptance of the norm-based model is grounded in a simple theoretical distinction between the qualitative predictions of norm- and exemplar-based models of face adaptation. Namely, exemplar-based accounts of face adaptation are thought to predict that face aftereffects will produce a local centred reduction in exemplar activation in the region surrounding the adaptor. Resultantly, following adaptation, the perception of a newly presented face will be biased away from the adaptor location, with the direction and magnitude of an aftereffect depending on the location of the target face relative to the adaptor. In contrast, norm-based accounts of face adaptation generally explain aftereffects in terms of a change in the location of the norm relative to which all faces are encoded, with the magnitude and direction of the aftereffect depending on the position of the adaptor relative to the norm (e.g. Leopold et al., 2001; Rhodes & Jeffery, 2006; Robbins et al., 2007). Specifically, adaptation in a norm-based model causes a bias in perception in a direction that is opposite to the way that the adaptor deviates from the norm.

While the weight of opinion is clearly heavily in favour of a norm-based account, there are actually relatively few distinct paradigms in the face adaptation literature that are thought to directly address the norm-versus-exemplar debate. For example, Leopold et al. (2001) reported a paradigm in which anti-face adaptors were used to bias the perception of morph faces lying on a trajectory between the average face and an opposite target, providing a putative demonstration that adaptation biases perception towards an opposite face (reviewed in Section 1.4.1, see Figure 4.2a). However, although this paradigm was originally taken as evidence of norm-based coding it has since been pointed out that the findings would also be consistent with an exemplar-based account (Rhodes & Jeffery, 2006; Rhodes et al., 2005; Tsao & Freiwald, 2006). Specifically, the facilitatory effect that adaptation to an anti-face has on the identification of the target could either be explained as a general bias away from the adaptor location (exemplar account) or as a bias in perception towards the opposite face (norm account). Nevertheless, many studies have used this paradigm to compare adaptive coding mechanisms between children and adults, using the results to support the conclusion that children have norm-coded face representations (Nishimura et al., 2008; Nishimura et al., 2011; Pimperton et al., 2009). Additionally, and to the same end, the paradigm has been used to investigate the face-space of adults with prosopagnosia (Nishamura et al., 2010; Palermo et al., 2011).

A second, related, paradigm, introduced by Rhodes & Jeffery (2006), is often cited as providing a more direct measure of the direction of aftereffects in face-space (e.g. Rhodes & Jeffery; Rhodes & Leopold, 2011; Rhodes et al., 2005; Tsao & Freiwald, 2006). This paradigm (reviewed in Section 1.4.1) aims to more clearly disentangle the 'general bias' predicted by an exemplar-based model from the 'opposite bias' predicted by the norm-based model. To do this, Rhodes and Jeffery introduced non-opposite morph trajectories into the paradigm used by Leopold et al. (see Figure 4.10a). In line with a norm-based model, they reported that aftereffects were stronger when measured along the opposite trajectory than they are when measured along the non-opposite trajectory Indeed, this pattern of results has now been demonstrated in children (Jeffery et al., 2011) and provides general support for the claim that adaptation biases perception in a direction that is opposite to the deviation of the adaptor face from the norm.

Perhaps the strongest evidence that face aftereffects cannot be explained by a general bias away from the adaptor location was provided by Robbins et al. (2007). In their study (reviewed in Section 1.4.1), they created a continuum of faces between a face with both eyes raised as far as the hairline (raised 50 pixels) and an undistorted target face (eyes raised 0 pixels). By measuring the perceived similarity of faces along the continuum to the target face both before and after adaptation to an intermediate face (eyes raised 20 pixels) they were able to demonstrate that adaptation had moved the perception of all faces on the continuum in the direction of the target. Indeed, this is what would be expected if an aftereffect biased perception in the opposite direction to the adaptor face (as in norm-based accounts). In contrast, if adaptation was a general bias away from a locally adapted region (as in exemplar-based accounts) then faces that were more extreme than the adaptor (eyes raised more than 20 pixels) ought to have been biased away from the target.

Overall then, it would seem that the evidence does indeed favour an interpretation of face aftereffects in which they are described as a bias in an opposite direction to the adaptor rather than as a general bias away from the adaptor location. The final line of evidence, which has been used to support norm-based coding in a number of studies (e.g. Jeffery et al., 2010; Skinner & Benton, 2010; Susilo et al., 2010a; Susilo et al., 2010b), is based on demonstrating that aftereffect magnitude increases with the distance of the adaptor face from the norm. For example, Leopold and Bondar (2005) demonstrated that when the anti-face used in the Leopold et al. (2001) paradigm is exchanged for an average face the facilitatory effect of adaptation on the identification of the target face is almost eliminated. This finding is consistent with the idea that more distinctive faces should cause a bigger change in the perceived location of the norm, with an average face causing the norm to change very little, if at all. Indeed, Susilo et al. (2010b) further demonstrated that the magnitude of face aftereffects increases almost linearly as the adaptor is moved further from the average face.

Despite the convergence of the literature on a norm-based model of face recognition, theoretical accounts of face adaptation in norm- and exemplar-based models remain untested through computational instantiation. Indeed, while verbal theorising and thought experiments undoubtedly provide an insight into the predictions of such models, the effects of high dimensionality on exemplar density and the non-linearity of activation functions can make the behaviour of such models difficult to intuit (Burton & Vokey, 1998; Lewis, 2004; Hintzman, 1990). Furthermore, previous research into the effect of caricature and distinctiveness on face recognition has revealed some significant similarities between the predictions of norm- and exemplar-based models that ought at least to suggest caution before declaring the debate to be conclusively settled. Thus, in the following sections several possible face-space implementations of face adaptation will be explicitly described and instantiated, testing their predictions across the range of adaptation paradigms just discussed.

4.2 Implementing Models of Face Adaptation

To investigate the qualitative and quantitative predictions of norm- and exemplar-based models on a range of face adaptation paradigms, a simple two-layer face-space model was implemented. The first layer of the model, the face-space, consisted of a population of face units representing the statistics of a given population of faces. The response properties of the face units were dependent on the model under consideration; specifically, three accounts were investigated, a traditional norm-based model (e.g. Giese and Leopold, 2005, see Section 1.2.1), a two-pool norm-based model (e.g. Rhodes & Jeffery, 2006, see Section 1.5), and an exemplar-based model (e.g. Lewis, 2004, see Section 1.2.2). Adaptation was implemented in the face-space such that the post adaptation response of a given unit was inversely proportional to its degree of activation by the adapting stimulus. The second layer of the model, the decision layer, was a linear neural network trained on the outputs of the face-space to reflect the specific demands of a given adaptation paradigm. Importantly, the implementation of the decision layer was identical for all three versions of the model.

4.2.1 Modelling the Face-Space Layer

For every simulation of the model, a unique population of 500 face units[4] was created, with the location of a given unit randomly selected from a normal distribution (SD=1.0) along each of the face-space dimensions. Thus, the faces were assumed to come from a single homogenous population (e.g. White male faces). The number of dimensions n was varied between 2 and 50, a range that includes the most likely range of face-space dimensions suggested by Lewis (2004).

4.2.1.1 The Exemplar-Based Model

In the exemplar-based model, the activation of a face unit **k** in response to a target face **f** was determined by a Gaussian function of the distance between the

---

[4] Testing the model using 250 and 1000 face units revealed that the number of face units had little influence on the qualitative predictions of the models.

target face and the face unit. In addition, the free parameter η controlled the broadness of tuning for each Gaussian function (see Figure 4.1a):

$$act_{k,f} = \exp\,(-\frac{|\mathbf{k}-\mathbf{f}|^2}{2\eta^2}) \qquad (1)$$

4.2.1.2 The Traditional Norm-Based Model

In the traditional norm-based model, the activation of face unit **k**, in response to the presentation of a target face **f** was proportional to the cosine (normalised to be between 0 and 1) of the angle, relative to the norm **m**, between the face unit and the target face scaled by distance of the target face from the norm (see, Giese & Leopold, 2005). In Addition, the free parameter v represented the broadness of tuning for each norm-coded face unit (see Figure 4.1b):

$$act_{k,f} = \ |\mathbf{f}-\mathbf{m}|(\frac{(\mathbf{f}-\mathbf{m})(\mathbf{k}-\mathbf{m})}{2|\mathbf{f}-\mathbf{m}||\mathbf{k}-\boldsymbol{m}|}+\frac{1}{2})^{v} \qquad (2)$$

4.2.1.3 The Two-pool Model

The two-pool norm-based model was implemented in line with descriptions in the literature (e.g. Rhodes & Jeffery, 2006). To formalise the model one can assume that for each face unit, there is an opposing unit, positioned on the opposite side of the norm (while this is not actually necessary it is a convenient way to formalise the two-pool model, see Chapter 1). Thus, the location of the norm is implicitly represented in the proportional activation of a pair of opponent coded face units, whereby, the activation function for each individual unit is a Gaussian function, identical to that in the exemplar-based model (see

Equation 1). To implement this, for a given face unit $\mathbf{k_i}$there was an opposite unit $\mathbf{k_2}$ created that lay on the opposite side of the norm. The activation of a given unit in the model was given by dividing its activation by the sum of the activation of both face units (see Figure 4.1c):

$$act_{k,n} = \frac{act_{k,n}}{\sum_{n=1}^{2} act_{k,n}} \quad (3)$$

Whereby, n can take values of 1 or 2 for opposing pools.

4.2.2 Modelling the Decision Layer

The decision layer was identical in all three models, consisting of a single-layer neural network with a linear transfer-function and a set of output nodes corresponding to the p target identities to be learned in a given adaptation paradigm. On each simulation of the model, the network was trained on the activation vector from the units in the face-space layer in response to the target identities. The learning rate was set using the Widrow-Hoff learning algorithm and continued to an error of 0.01 or 1000 epochs. Additionally, the response for each node was set to give 1.0 for an exact match to one of the learned target identities and -1.0 for all other learned identities.

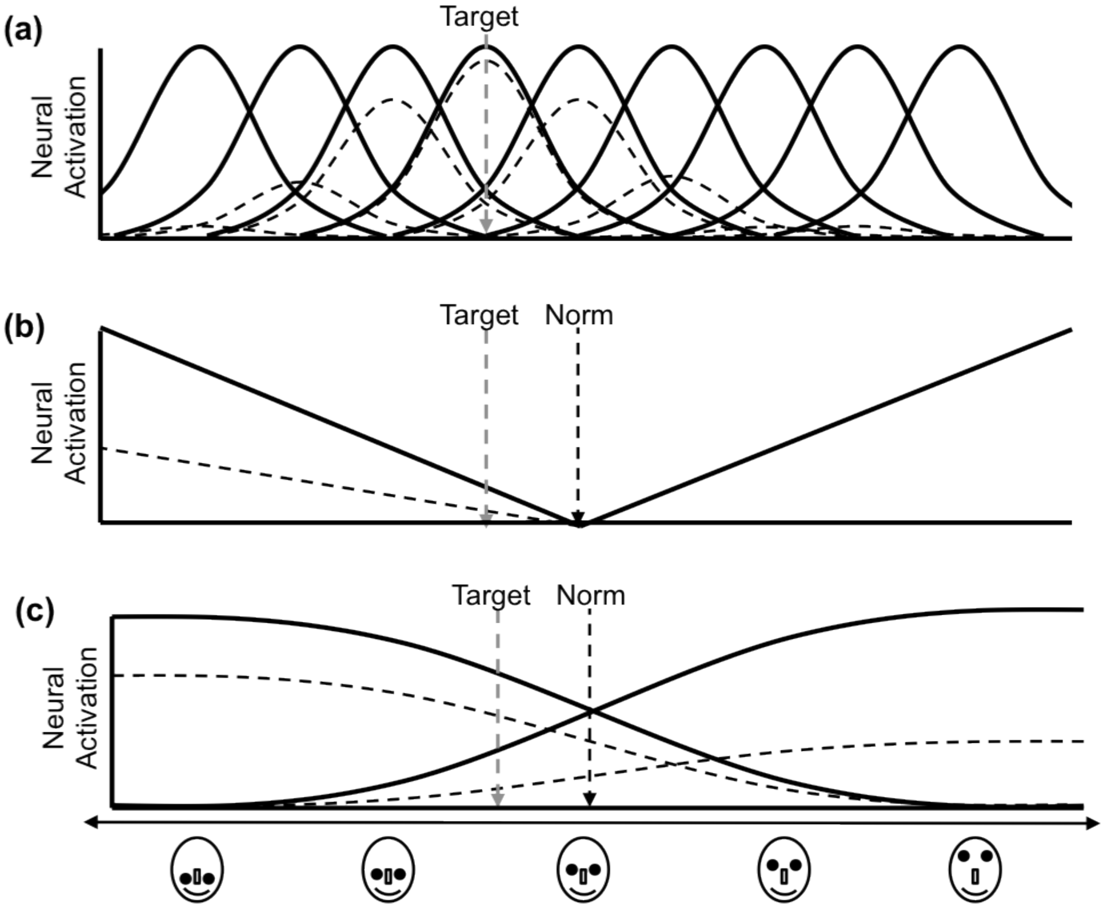


Figure 4.1. Representation of the activation in the three models implemented. The solid line represents the activation potentials of the face representations in each model. The dotted lines represent the activation produced by the target face. (**a**) Exemplar-based model: activation was a Gaussian function of the Euclidean distance between a representation and the target (**b**) Traditional norm-based model: activation was a function of the angle between a representation and the target, with the activation strength scaled by the distance of the target from the norm (**c**) Two-pool norm-based model: activation of the two pools was a function of distance as in the exemplar-based model. However, as activation was normalised against an anti-face it was also related to direction.

4.2.3 Modelling Face Adaptation

The post-adaptation response of a face unit in the exemplar and two-pool models was inversely proportional to the activation of the face unit in response to the adaptor face (e.g. Rhodes & Jeffery, 2006; Robbins et al., 2007). Specifically, the post-adaptation activation of a face unit was the product of its pre-adaptation activation to the target and an inhibition factor $\left(1 - \alpha . act_{k,a}\right)$ that was inversely proportional to the activation of the face unit to the adaptor face **a**. Additionally, the free parameter α controlled the degree of adaptation:

$$act2_{k,f} = act_{k,f}\left(1 - \alpha \, . \, act_{k,a}\right) \tag{4}$$

Adaptation in the traditional norm-based model was implemented in a similar manner except that, as the activation of face unit is not constrained to a number between zero and one the inhibition factor was normalised as follows:

$$act2_{k,f} = act_{k,f}\left(1 - exp^{-\theta . \ \alpha \, . act_{k,a}}\right) \tag{5}$$

Whereby, the parameter θ controls the shape of the inhibition function.

4.2.4 Overview of Simulations

To reiterate, the aim of the experiments reported in this chapter was to investigate the qualitative and quantitative predictions of the three different versions of face-space (traditional norm-based, two-pool norm-based and exemplar-based) across a number of putatively diagnostic face adaptation paradigms. As reviewed in the introduction, there are actually relatively few distinct paradigms that have been used to differentiate norm and exemplar-based models with different studies broadly tapping into one of two related theoretical claims about norm-based adaptation. Namely, whether adaptation biases perception in an opposite direction from the direction that the adaptor deviates from the norm (Leopold et al., 2001; Rhodes & Jeffery, 2006; Robbins et al., 2007) and whether adaptation increases with the distance of the adaptor from the norm (Leopold & Bondar, 2005; Susilo et al., 2010b).

To this end, the approach taken here was to focus on the qualitative predictions of the models across as broad a range of parameter values as possible. This was achieved by changing the free-parameters of each model in small increments, recording the qualitative pattern of results for each parameter set on each paradigm, and creating a 'qualitative map' of each model's predictions (parameter values are given in Table 1). The reason for this approach was two-fold. First, for all of the paradigms being investigated it is the qualitative pattern of results that is important in differentiating norm- and exemplar-based accounts. Second, testing the models over a very broad range of parameters, rather than simply finding a best fitting parameter set will provide a fuller understanding of the conditions in which each model is able to fit the data. However, in addition to qualitatively mapping the parameter space, a quantitative 'best-fit' was also determined for each model on each paradigm by comparing the fit for each parameter set to data taken from the literature. It should be noted that the point of this best-fit is not to compare across models, but rather to provide some confirmation of the model fitting.

Parameter	Lower bound	Upper bound	Step size
η	0.6	12	0.3
$\boldsymbol{v}$	0.6	12	0.3
$\boldsymbol{a}$	0.2	0.8	0.2
$\boldsymbol{\theta}$	0.5	1.5	0.1

Table 1. Parameter values used in Experiment 5a, Experiment 6 and Experiment 7a

4.3 Experiment 5a: Modelling Leopold et al. (2001)

In Experiment 5a, a version of the paradigm used by Leopold et al. (2001) was instantiated in each of the three models. This paradigm has been widely used in the literature to support claims about norm-based coding (e.g. Leopold et al.; Nishamura et al., 2010; Nishimura et al., 2008; Nishimura et al., 2011; Palermo et al., 2011; Pimperton et al., 2009). Indeed, in line with the predictions of a norm-based model, Leopold et al. demonstrated that the identification of a target face, say, Adam, was facilitated by exposing participants to the opposite face, anti-Adam (see Figure 4.2). In contrast, adaptation to a non-opposite anti-face generally appeared not to facilitate the recognition of the target face; indeed, it may even hinder identification.

While this study is frequently cited as providing evidence for a norm-based model it has been pointed out that the results are also consistent with an exemplar-based model (e.g. Rhodes & Jeffery, 2006). Specifically, in an exemplar-based model the result could be explained as a bias away from the adaptor location, rather than as a bias towards the opposite face. With this in mind, Experiment 5a aimed to provide a starting point for a computational exploration of adaptation aftereffects in norm- and exemplar-based models. Indeed, although the exemplar-based model might be expected to make predictions that are qualitatively consistent with the findings of Leopold et al. (2001) this prediction remains to be tested, as do the predictions of both the traditional and two-pool norm-based models.

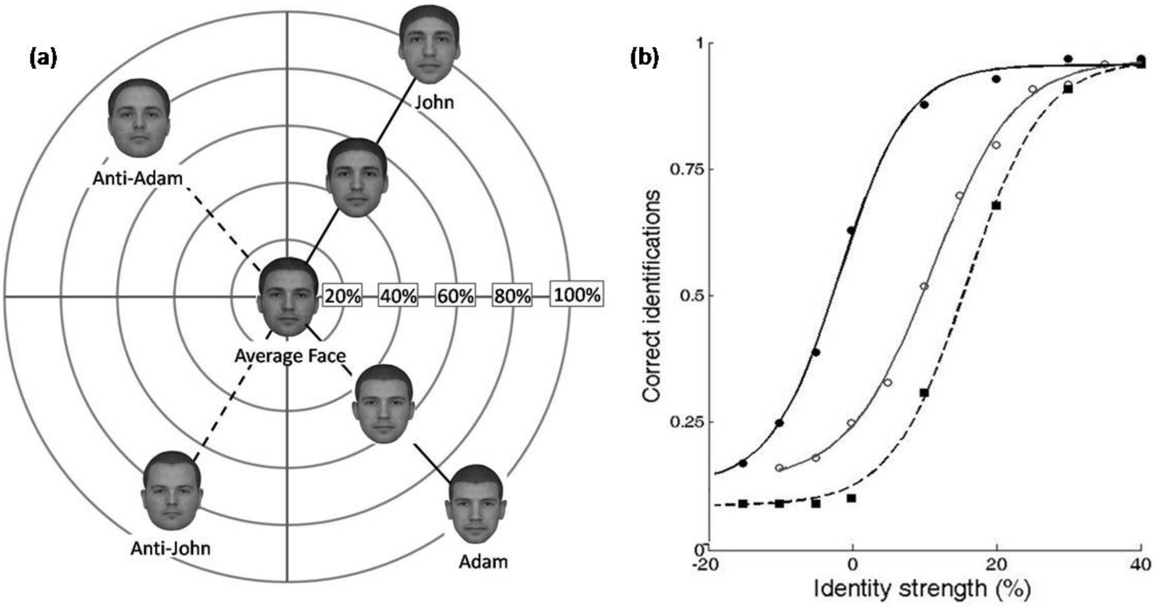


Figure 4.2: **(a)** Schematic representation of the stimuli used by Leopold et al. (2001). All identity trajectories pass through the average face such that the 0% identity level of each target trajectory is equivalent to the average face. Anti-faces lay -80% opposite each of the corresponding faces. **(b)** Sensitivity to face identity with and without adaptation (data points taken from Leopold et al). Three conditions are shown: baseline responses (◦), responses following adaptation to an opposite anti-face (•), and responses following adaptation to a non-opposite anti-face (▪). The proportion of correct responses at each identity level has been averaged across the four identity trajectories, fitting a four parameter logistic function to the data.

4.3.1 Methods

4.3.1.1 Materials: Following Leopold et al. (2001), the face stimuli were generated in the context of a schematised face-space (see Figure 4.2a). Conceptually, all of the face stimuli lay on trajectories passing through the average face, with the average face defined by the origin of the face-space (rather than the average of the face unit locations in face-space). On each simulation of the model, four target faces were generated (Adam, Jim, John and Henry) selecting the position of each face along each of the face-space dimensions from a random normal distribution (SD=1.0). In addition, 8 jittered versions of each of the four target faces were created by adding a 5% noise vector, generated from a random normal distribution (SD=1.0), to each target face's position vector. Identity trajectories were constructed for each face by multiplying the position vector for each of the four target faces by a factor between 1 and -0.2 in 0.05 steps (4x25 faces). Thus, each identity trajectory extended from the 100% identity level (target face), through the average face, or 0% identity level, to the -20% identity level that lay on the opposite side of the average. Additionally, four anti-face adaptors were created, multiplying the position vectors for each of the target faces by -0.8, such that each adaptor corresponded to the -80% location on a given identity trajectory, opposite the corresponding target face.

4.3.1.2 Procedure: For each parameter set, for each model, 100 simulations were run. On each simulation, the decision layer was first trained on the output of the face-space layer in response to the 8 jittered versions of the four 100% target identities. Thus, there were four nodes in the decision layer, each corresponding to one of the four target faces. As the task was a forced choice task, the probability that the model would decide that a given test face f was one of the target faces k was given by the output of the node corresponding to the target face divided by the sum of the activation across all four nodes. Whereby, ϕ controlled the sharpness of the probability function and was set to vary between 1 and 40 in steps of 1:

$$P\left(D_k \middle| T_f\right) = \frac{exp^{\phi,actn_{k,f}}}{\sum_{i=1}^{4} exp^{\phi.actn_{i,f}}} \tag{6}$$

Three types of trial were run, Baseline trials, Opposite Adaptor trials, and Non-Opposite adaptor trials. Only the probability of the model making a correct response to a given target face was recorded. For example, the correct response for a given identity level of a given face, say 20% Adam, was considered to correspond to the target identity from which it was generated, in this case, Adam. On Baseline trials, the probability of the model correctly identifying each of the 25 identity levels on each of the four target trajectories was recorded. On Opposite Adaptor trials, the probability of the model correctly identifying each of the 25 identity levels on each of the four target trajectories following adaptation to a corresponding anti-face was recorded. For example, the probability of correctly identifying the 25 faces on the Bob target trajectory as Bob was recorded following adaptation of the face-space layer to anti-Bob. Finally, on Non-Opposite Adaptor trials, the probability of correctly identifying each of the 25 identity levels on each of the four target trajectories following adaptation to each of the three non-corresponding anti-faces was recorded.

4.3.2 Experiment 5a Results

The probability of the model making a correct response at each identity level on each of the three trial types, Baseline, Opposite Adaptation, and Non-Opposite Adaptation, was averaged across the four target trajectories. A four parameter logistic function was fitted to the data for each condition, for each simulation of the model. The effect of opposite and non-opposite adaptation was calculated relative to the baseline, taking the threshold at the inflection point (as in Leopold et al., 2001).

Two qualitative criteria were considered. First, the value for the opposite adaptation should be positive, indicating that adaptation facilitated the correct identification of the target faces. Second, the non-opposite adaptation ought to facilitate recognition less than opposite adaptation. To this end a two-tailed t-test was conducted on the data with the criterion for significance set to $p<0.01$. Parameter sets for which both criteria were fulfilled were scored as qualitatively correct, parameter sets for which one or both of the criterion were significant but not in the expected direction were scored as incorrect, and parameter sets for which both criterion were not significant were scored as null. For simplicity, the data were then collated across values of ϕ (which determines the sharpness of the probability function for the decision layer) and θ (which determines the shape of the adaptation function in the traditional norm-based model) as follows: If there were any values of ϕ and θ that gave a correct qualitative result for a given parameter set then it was scored as qualitatively correct. If not, if there were any values of ϕ and θ that gave an incorrect qualitative result for a given parameter set then it was scored as qualitatively incorrect. Last, if there were no values of ϕ and θ that gave any significant results then it was scored as null. Note that, the data were collated in this manner because it was considered that values of ϕ and θ were not directly linked to the theoretical questions being addressed in the current model. However, to ensure that this method of collating the data did not influence the qualitative results the reverse procedure was also tested. That is to say, values of ϕ and θ that gave incorrect qualitative fits were given priority. This made no difference to any of the qualitative maps for any of the models. The results were then converted into qualitative maps of the parameter space (see Figure 4.3a, 4.4a, & 4.5a). An examination of the qualitative maps revealed that all three models made accurate qualitative predictions across a broad range of parameters. Indeed, the exemplar-based model and the two-pool norm-based model made no incorrect qualitative predictions for any parameter values.

An approximate quantitative best-fit was also calculated, fitting a four-parameter logistic function to the average data across participants on each parameter set, and calculating the SSE between these points and the corresponding points in the data taken from Leopold et al. (2001). Thus, the data presented in Figures 4.3b, 4.4b, & 4.5b correspond to the best fits obtained from each model. Note that only the qualitative map corresponding to the best-fit value of α (which controls the degree of adaptation) is shown, though there was very little difference between the qualitative maps for a given model. An examination of the quantitative fits revealed that in line with the results found by Leopold et al. (2001), adaptation to an opposite anti-face facilitated identification of the target in all three models (indicated by a shift in the logistic function to the left of baseline). Moreover, adaptation to a non-opposite anti-face slightly hindered identification of the target (indicated by a shift in the logistic function to the right of the baseline).

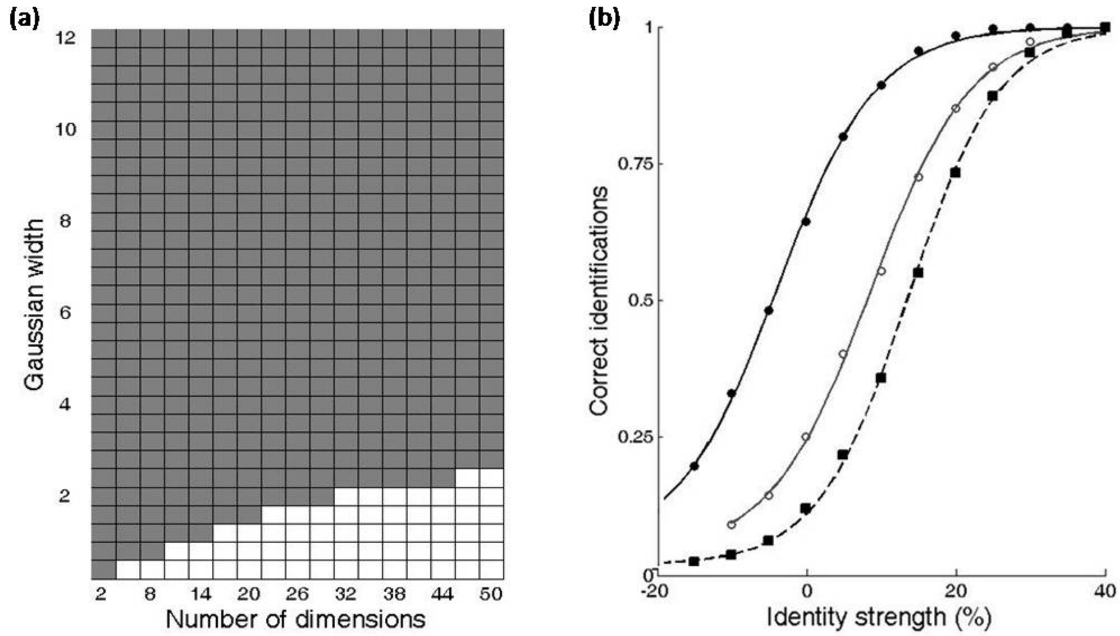


Figure 4.3. Simulation results from the exemplar-based model. **(a)** Map of the parameter space explored in Experiment 1a showing the qualitative fit with the findings reported by Leopold et al. (2001). Grey sqares (■) represent parameter values on which the model made qualitatively accurate predictions, white squares (□) represent parameter values that produced no signifcant adaptation (null), and black squares (■) represent parameter values in which the model made predictions that were qualitatively inconsistent with the findings of Leopold et al. **(b)** Quantitative best-fit to the data from Leopold et al. Three conditions are shown: Baseline (○), Opposite Adaptation (●), and Non-Opposite Adaptation (■). The best-fit parameter values were found for n = 11, η = 5.2, α = 0.2, and ϕ = 22. Note, the qualitative map shown corresponds to the value of α for which the best-fit was obtained.

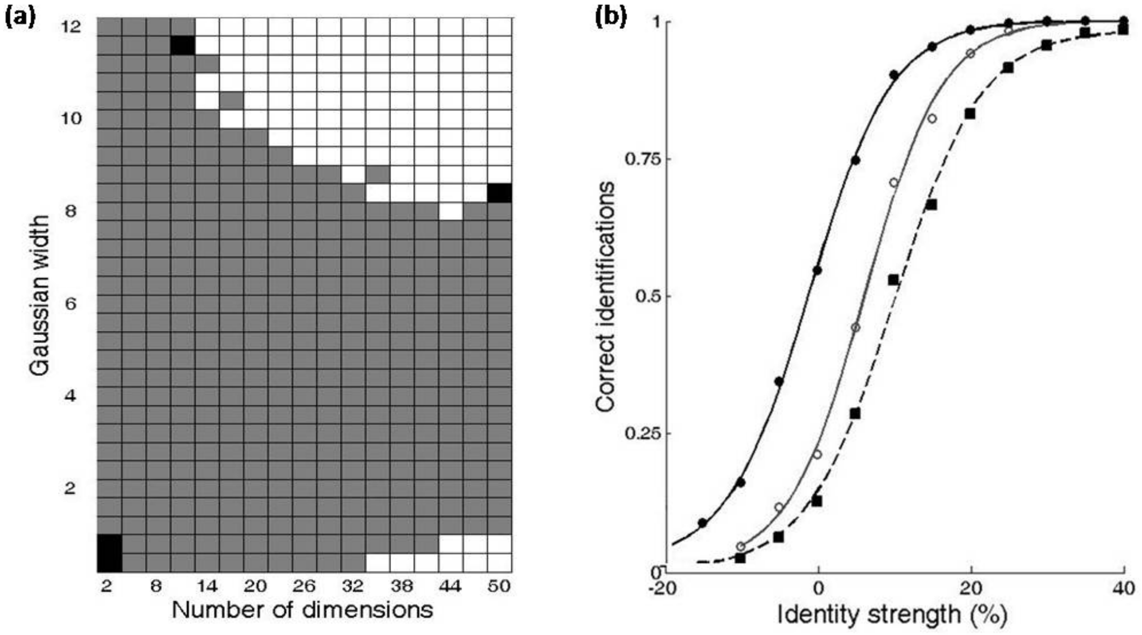


Figure 4.4. Simulation results from the traditional norm-based model (See Figure 4.3 legend for details). The best-fit parameter values were found for n = 8, v = 1.2, α = 0.8, ϕ = 30, and θ = *0.75.*

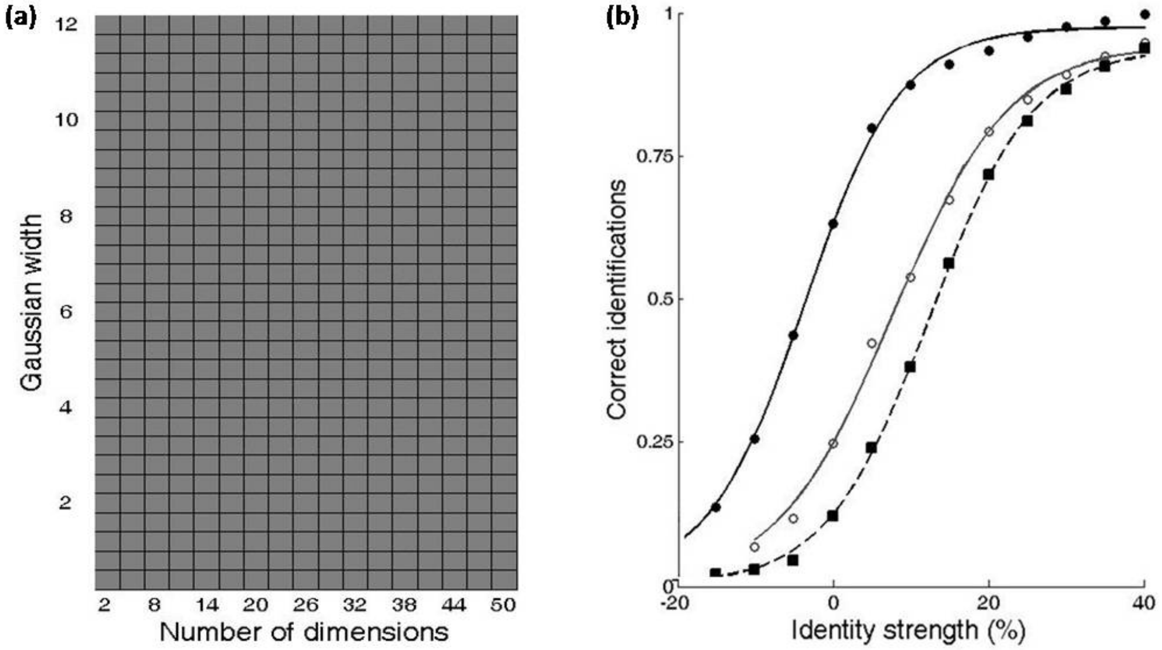


Figure 4.5. Simulation results from the two-pool norm-based model (See Figure 4.3 legend for details). The best-fit parameter values were found for n = 2, η = 0.8, α = 0.4, and ϕ = 22.

4.3.3 Experiment 5a Discussion

In Experiment 5a, three versions of face space, an exemplar-based face-space, a traditional norm-based face-space, and a two-pool norm-based face-space, were instantiated in order to investigate their qualitative predictions on the adaptation paradigm used by Leopold et al. (2001). All three models gave good qualitative fits across a broad range of parameters, indicating that adaptation to opposite adaptors significantly facilitated identification. Furthermore, adaptation to opposite adaptors facilitated recognition more than adaptation to a non-opposite adaptor. Indeed, in line with the data of Leopold et al., the best-fit figures indicate that adaptation to a non-opposite face may actually hinder recognition when compared to no adaptation at all. However, this was not specifically tested for in the qualitative maps, as there was no strong theoretical reason to suspect that it should always be detrimental to adapt to a non-opposite face.

Although the original study by Leopold et al. is often cited as support for a norm-based model of face recognition (e.g. Leopold et al.; Nishamura et al., 2010; Nishimura et al., 2008; Nishimura et al., 2011; Palermo et al., 2011; Pimperton et al., 2009), it is not entirely surprising that all three versions of the model provided good qualitative fits. Indeed, as several authors have pointed out, these results are consistent with the current understanding of adaptation in an exemplar-based model in which adaptation biases perception away from the adaptor location in all directions (e.g. Rhodes & Jeffery, 2006; Tsao & Freiwald, 2006). Nonetheless, the models implemented here provide the first attempt to actually test these theoretical predictions, providing a useful conformation of previous claims in the literature.

4.4 Experiment 5b: Modelling Leopold & Bondar (2005)

The results of Experiment 5a support the view that the paradigm employed by Leopold et al. (2001) is not sufficient to differentiate norm- and exemplar-based accounts of face-space. Thus, in Experiment 5b, a simple extension to the paradigm was implemented, exchanging the anti-faces used in Experiment 5a for an average face and implementing each of the models using the best-fit parameter values found in Experiment 5a.

Norm-based accounts of adaptation suggest that the magnitude of adaptation produced by an adapting stimulus is proportional to its distance from the norm. Thus, if the adaptor face is an average face then a norm-based model ought to predict that there will be no observable aftereffect. In contrast, exemplar-based accounts of adaptation suggest that the magnitude of adaptation produced by an adapting stimulus is dependent on the location of the subsequently presented target face (see Figure 4.6). Thus, if the adaptor face is an average face then an exemplar-based model should predict that there would still be a strong aftereffect biasing perception towards the target face.

Using this paradigm Leopold and Bondar (2005) found that, whereas adaptation to -80% anti-faces (such as those used in Experiment 5a) produced a marked adaptation aftereffect, adaptation to a 0% average face produced no observable aftereffect. These results were interpreted to be inline with a norm-based model and contrary to the predictions of an exemplar-based model. Thus, Experiment 5b provides a direct test of predictions of norm- and exemplar-based models in relation to the theoretical claims made in the literature.

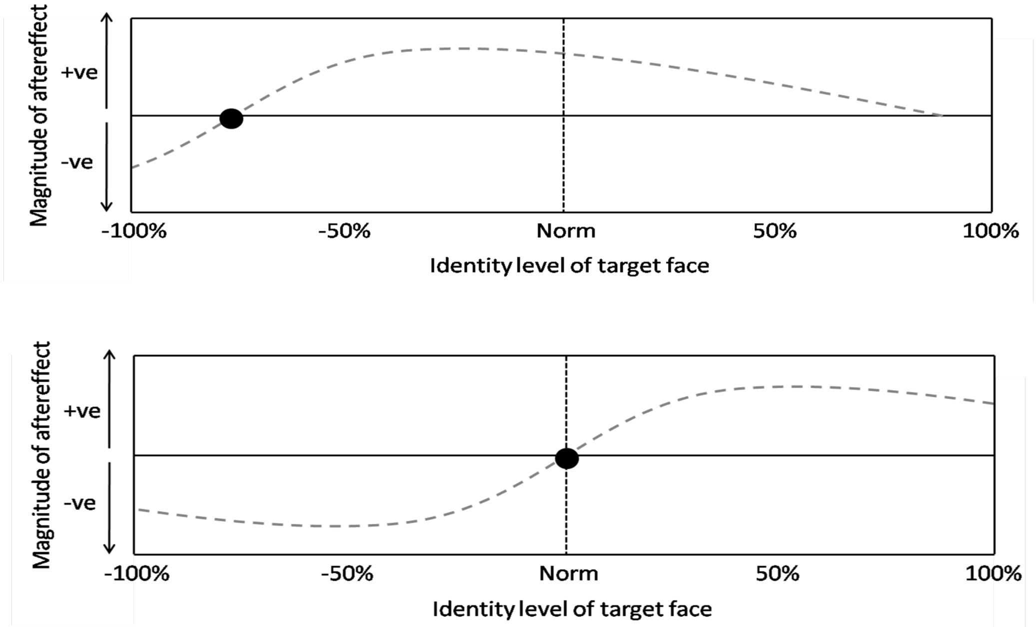


Figure 4.6. Hypothetical magnitude and direction of adaptation aftereffects (- -), resulting from adaptation to different adaptors (●), versus Identity level of target face in an exemplar-based model. **Top:** Adaptation to a -80% anti-face (as used in Experiment 5a). **Bottom*:*** Adaptation to a 0% average face. Assuming that the target trajectory extends between 100% identity level and -20% identity, as in Experiment 5a & 5b, adaptation to a –80% face ought to lead to a general bias of faces on the target trajectory towards the 100% target. However, adaptation to an average face ought to bias perception of faces on the target trajectory even more strongly towards the 100% target.

4.4.1 Method

4.4.1.1Materials & Procedure: For each model 100 simulations were run using the best-fit parameter sets from Experiment 5a. The stimuli used in this study were created in an identical manner to those described in Experiment 5a except that the -80% anti-face was replaced by an average face adaptor. Thus, in this experiment there were only two conditions: Baseline, and Average Adaptation. Note that, there is no Non-Opposite Adaptation because all adaptors for all trajectories were always at the same location (That is to say 0% average face). As with Experiment 5a, the average face was taken to correspond to the origin of the space.

4.4.2 Experiment 5b Results

The probability of the model making a correct response at each identity level on the two trial types, Baseline, and Average Adaptation, was averaged across the four target trajectories. A four parameter logistic function was fitted to the data for each condition, for each simulation of the model. Logistic functions were also fitted to the mean data across simulations revealing that there was very little aftereffect produced by adaptation to an average face in any of the three models (Figure 4.7). This observation was confirmed by a comparison of the thresholds at the inflection point for Average Adaptation to the Baseline threshold for each fit on each simulation of a given model (as in Leopold & Bondar, 2005) with two-tailed t-tests revealing that Average Adaptation condition did not significantly differ from Baseline condition in any of the three models $p>0.01$.

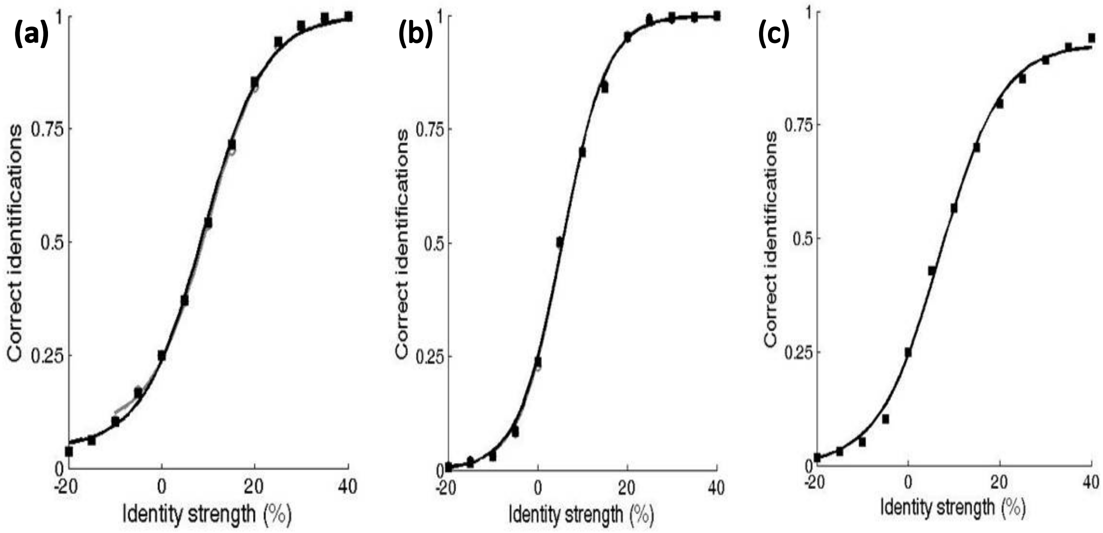


Figure 4.7. Sensitivity to face identity with and without adaptation to the average face (Experiment 5b) for the exemplar-based model **(a)**. traditional norm-based model **(b)**, and two-pool norm-based model **(c)**. Two conditions are shown: Baseline (○), and Average Adaptation (●). The proportion of correct responses at each identity level has been averaged across the four identity trajectories, fitting a four parameter logistic function to the data.

4.4.3 Experiment 5b Discussion

In Experiment 5b, the predictions of the three models on a modified version of the task used by Leopold and Bondar (2005) were tested. Specifically, whereas Experiment 5a investigated the effect that adapting to a -80% anti-face had on sensitivity to identity, Experiment 5b investigated the effect that adapting to an average face had on sensitivity to identity. The results indicated that adaptation to an average face, corresponding to the 0% identity level, had very little effect on sensitivity to identity in any of the three models.

These findings are in line with a host of findings in the literature that suggest that adaptation to the average face does not result in a significant aftereffect (e.g. Jeffery et al., 2010; Leopold & Bondar, 2005; Robins et al., 2007; Skinner & Benton, 2010; Susilo et al., 2010a; Susilo et al., 2010b; Webster & MacLin, 1999). However, all of the aforementioned studies have used this finding as evidence in favour of a norm-based model of face-space. It is, therefore, surprising that the results of Experiment 5b indicate that both the two norm-based models and the exemplar-based model made predictions that were in-line with the empirical data. This provides a substantial challenge to the theoretical claims that adaptation in an exemplar-based model is simply a bias away from the location of the adaptor. Indeed, if this had been the case then it might have been expected that adaptation to the average face should actually facilitate the identification of the target faces more than adaptation to the anti-face (see Figure 4.6).

One possible issue is that the paradigms used in Experiment 5a and 5b involve making a forced choice between four possible identities (implemented in the decision layer). It might be suggested that it is the nature of this task, rather than the nature of the face-space, that caused adaptation to appear centred on the average face. For example, the fact that the identity trajectories of the four target identities all pass through the centre of the space might have meant that adaptation to the average did not specifically favour the identification of any of the four targets. One way around this is to measure the direction of the aftereffects directly in the face-space layer; this is investigated in Experiment 5c.

4.5 Experiment 5c: Adaptation in the Face-Space Layer

In Experiment 5c, the predictions of the exemplar-based model were further tested using the best-fit parameters from Experiment 5a. Rather than recording the responses of each model in the decision layer the 'centre of gravity' of activation (COG) was measured in the face-space layer. The COG has previously been used as a measure of the perceived location of a stimulus in single dimensional exemplar coding models of low- and mid-level vision (Suzuki, 2005).

The series of simulations carried out in Experiment 5c were based on predictions made by Suzuki (2005) designed to differentiate single dimensional exemplar models of aspect ratio encoding from two-pool models. The idea is to measure the perceived location of a particular stimulus on a given dimension (COG) before and after adaptation to a range of adaptors lying at different points along that dimension. By plotting the perceived location of the target on the dimension after adaptation to each adaptor a curve can be obtained (see Figure 4.8). Suzuki predicted that in the exemplar-based model the adaptor tuning curves would be centred on the perceived target location before adaptation. In other words, when the adaptor and target face were identical there would be no bias in perception. This is because the adaptor would adapt nearby exemplars biasing perception away from that location equally in all directions. Thus, if the adaptor and target location were the same then the target would not be biased in any particular direction more than others. In contrast, Suzuki predicted that in a two-pool model the adaptation tuning functions would be centred on the average face because adaptation to the average face would adapt both pools equally resulting in no change in the perceived location of the target.

While theoretical coding models of low- and mid-level vision are outside of the scope of this thesis what is of interest is the relationship between these predictions and the current norm-versus-exemplar debate. Indeed, this paradigm directly addresses some of the fundamental differences thought to differentiate norm and exemplar models of face adaptation. Namely, it clarifies the relationship between the target and the adaptor, which is important in an exemplar model, and it clarifies the relationship between the adaptor and the norm, which is thought to be important in a norm-based model.

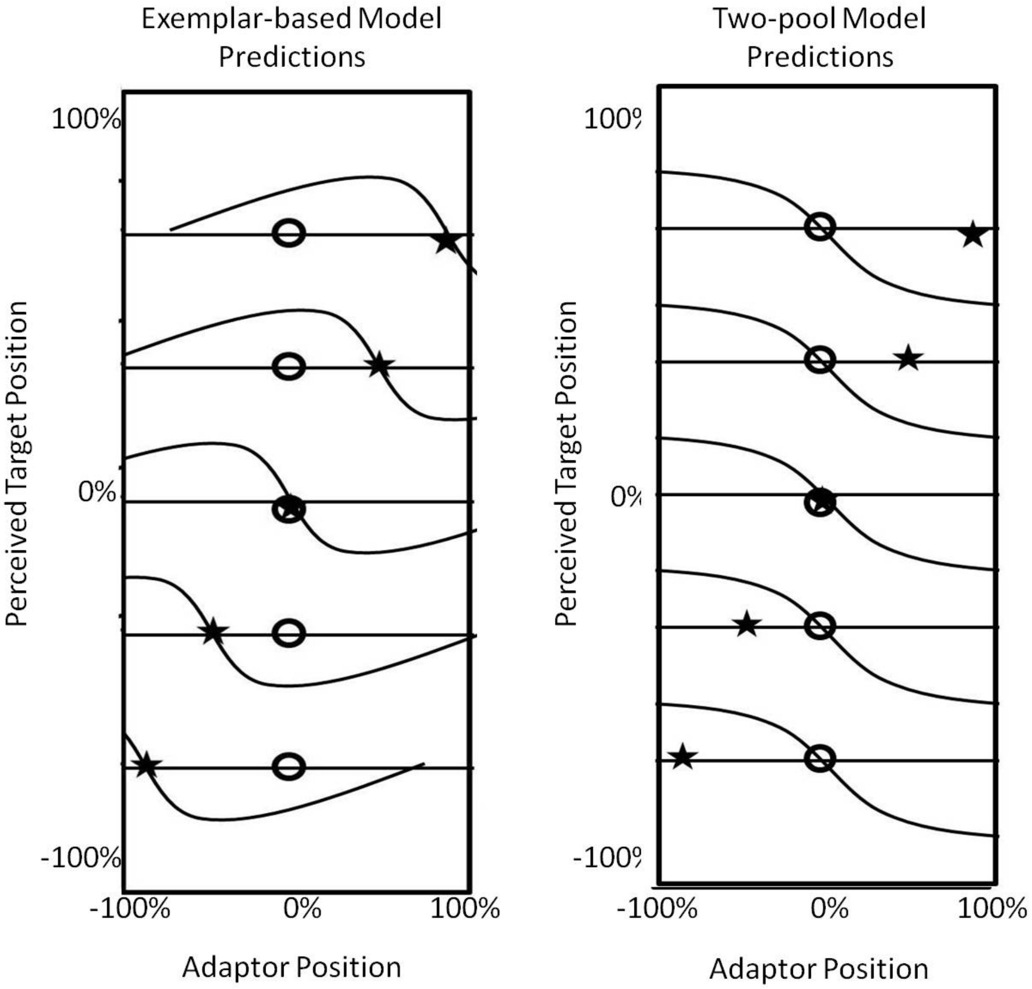


Figure 4.8. Expected tuning curves of aftereffects in an exemplar-based model (left), and a two-pool norm-based model (right), adapted from Suzuki (2005). The perceived location of the target face before adaptation is represented by the filled black star (★) and the location of the norm, or average face, is represented by the open circle (○). The solid black curves indicate the perceived target location following adaptation to adaptors at different identity levels. The tuning curves from the exemplar-based model are predicted to pass through the target location, indicating that perception is biased away from this point. However, the tuning curves from the two-pool model are expected to always pass through the norm.

4.5.1 Method

4.5.1.1 Materials: For each simulation of the exemplar-based model, a single face was generated, selecting its locations along each of n-dimensions from a random normal distribution (SD=1.0). An identity trajectory was constructed by multiplying the position vector of the target face by a factor between -1 and 1 in 0.05 steps (41 faces). Thus, the identity trajectory extended from the -100% identity level, through the average face, or 0% identity level, to the 100% identity level that lay on the opposite side of the average. Target faces were chosen to be the -100%, -50%, 0%, 50% and 100% identity levels, while all 41 identity levels served as adaptors.

4.5.1.2 Procedure: *Pre-Adaptation* - One hundred simulations of the exemplar-based model were run. On each simulation, the centre of gravity (COG, See Equation 7) of activation in the face-space layer was determined for each of the five target faces. The position of the centre of gravity was then projected onto the identity trajectory to determine the component of the activation in the direction of the measured trajectory.

$$COG = \frac{\sum_{i=1}^{n}(x_i.Act_i)}{\sum_{i=1}^{n} Act_i} \qquad (7)$$

Post-Adaptation- The activation of each of the 41 adaptor faces was then obtained and used to calculate the post adaptation activation of each of the five targets (5x41 conditions). Finally, the COG for each of the five targets after adaptation to the 41 adaptors was projected onto the identity trajectory.

4.5.2 Experiment 5c Results

For each of the five target-faces (-100%, -50%, 0%, 50%, and 100%) the location of COG on the identity trajectory before adaptation was averaged across simulations. The location of each of the five target-faces after adaptation to each of the 41 adaptor levels was also averaged across simulations and the results plotted against the identity level of the adaptor (Figure 4.9). There appeared to be no particular relationship between the target location and the adaptor tuning function. Rather, the adaptor tuning function was approximately centred on the average face, indicating that adaptation to the average face caused no change in the perceived location of the target on the identity trajectory. Moreover, as the difference between the perceived location of the target before and after adaptation is greater for adaptors that are further from the average face, it would appear that the magnitude of adaptation increases with the distance of the adaptor from the average face.

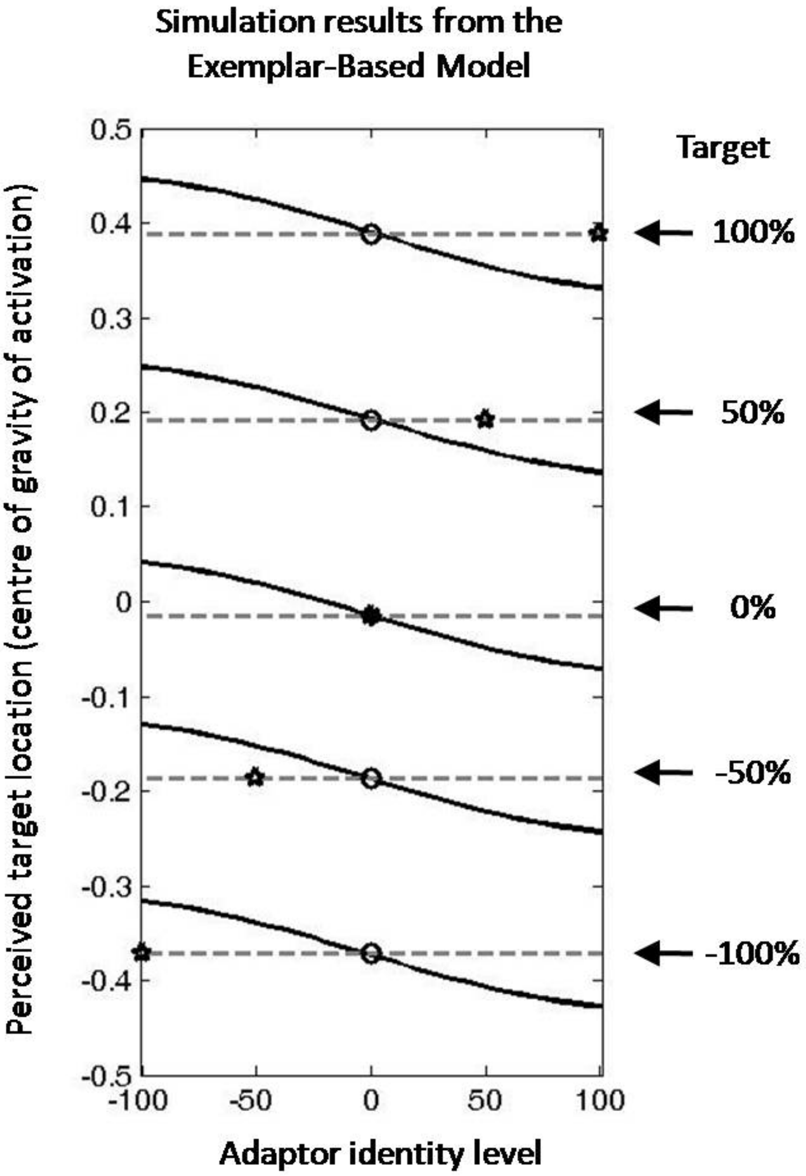


Figure 4.9. Tuning curves of aftereffects from the exemplar-based model in Experiment 5c. Perceived target location (COG) versus Adaptor identity level. The location of the target face is represented by the filled black star (★) and the location of the norm is represented by the open circle (○). The solid black curves indicate the perceived target location following adaptation to adaptors at different identity levels. The horizontal dashed lines represent the perceived target location.

4.5.3 Experiment 5c Discussion

In Experiment 5c, the predictions of the exemplar-based model were further investigated using the best-fit parameters from Experiment 5a to probe the shape of the adaptor tuning functions for targets at a range of different locations along a single identity trajectory. Comparing the tuning functions obtained from the model (Figure 4.9) against the tuning functions predicted by Suzuki (2005) for an exemplar or multi-channel coded model (Figure 4.8), it is clear that the predictions do not match with the model data. Specifically, whereas the predicted exemplar-based adaptation tuning functions pass through the target location, indicating a general bias in perception away from the adaptor location, the actual tuning functions from the exemplar-based model passed through, or very close to, the location of the average face. Thus, it appears, at least for this parameter set, that adaptation was centred on the average face with adaptation biasing perception towards a face with opposite attributes.

This finding is important, as it would suggest that exemplar-based models might be able to account for the range of aftereffects that have been reported in the literature. To investigate what properties of the model were responsible for this effect some additional simulations of the model were run. Surprisingly the results do not appear to be directly dependent on the dimensionality of the space or the density, that is to say, it is possible to obtain similar results using a single dimensional space with a square (rather than a normal) distribution of exemplars. However, one parameter that did make a substantial difference to the predictions was the broadness of tuning of each exemplar (this will be discussed more in the general discussion, Section 4.8). For, now it is enough to simply clarify that when the exemplars were narrowly tuned the aftereffects became more centred on the location of the adaptor, biasing perception away from this location. It is also worth noting that while the other parameters such as dimensionality or distribution type did not directly alter the predictions of the model they did interact with the broadness of tuning parameter. Specifically, as the number of dimensions increases the exemplars must be more broadly tuned. Similarly, if the distribution of exemplars is made squarer then exemplars must be more broadly tuned.

The results of Experiments 5a, 5b and 5c question prior assumptions regarding the range of results that should be considered as inconsistent with exemplar-based models of face-space and, thus, cast doubt on previous attempts to differentiate two-pool and exemplar based models. Therefore, Experiments 6 and 7 reconsidered results previously thought to be uniquely consistent with norm-based models of face-space.

4.6 Experiment 6: Modelling Rhodes & Jeffery (2006)

In Experiment 6, a version of the paradigm used by Rhodes and Jeffery (2006) was instantiated in each of the three models. This paradigm offers an extension to the paradigm introduced by Leopold et al. (2001) by allowing the direction of the observed aftereffects to be inferred. The reason that it is possible to infer the direction of aftereffects in the Rhodes and Jeffery paradigm is that both opposite trajectories, like those in Experiment 5a, and non-opposite trajectories are constructed. As with Leopold et al., four 100% target faces were constructed (Dan, Jim, Rob, and Ted) using the 100% faces to construct -80% opposite adaptors. However, unlike Leopold et al., non-opposite adaptors were also constructed to be equally far from the 100% targets as the -80% opposite adaptors. Importantly, the trajectory between the non-opposite adaptors and the 100% faces did not pass directly through the average (see Figure 4.10). Thus, if adaptation simply results in a biasing of perception away from the adapted region (as has been claimed for the exemplar-based model) then both opposite and non-opposite adaptors should facilitate recognition of the 100% target identity equally. However, if adaptation is centred on the average face then adaptation to an opposite adaptor should facilitate recognition more than adaptation to a non-opposite adaptor. This is because adaptation centred on the average face biases perception towards a face with opposite stimulus attributes.

Although this experiment was initially devised in order to distinguish between the predictions of two-pool norm-based models and exemplar-based models it would appear from Experiments 5a, 5b, and 5c that all three models may make qualitative predictions in line with those obtained by Rhodes and Jeffery (2006). Thus, Experiment 6 provides a further test of the predictions of the three models as well as giving an indication of the range of parameters over which the exemplar-based model predicts aftereffects that are average centred, as oppose to adaptor centred.

A final consideration is that in the experiment originally carried out by Rhodes & Jeffery they reported a substantial learning effect over the course of the adaptation trials, particularly for faces on the non-opposite trajectory (see Figures 4.10b and 4.10c). However, none of the theoretical models considered here make explicit predictions in this regard. Thus, for simplicity, no attempt was made to instantiate learning in the three models. Rather, the quantitative fitting was conducted using the post-learning baseline and adaptation data.

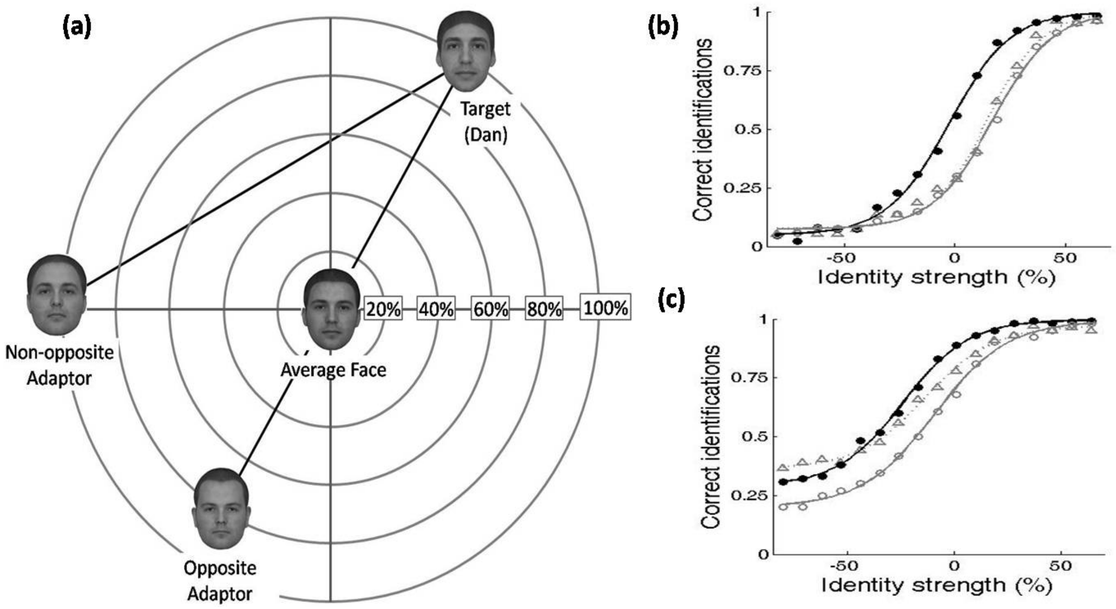


Figure 4.10. **(a)** Schematic representation of the stimuli used by Rhodes and Jeffery (2006). The opposite identity trajectories pass through the average face, whereas, the non-opposite trajectories do not. **(b)** Sensitivity to face identity with and without adaptation to an opposite face (data points taken from Rhodes and Jeffery). Three conditions are shown: baseline responses (○), post-learning baseline responses (△), and responses following adaptation to an opposite face (●). **(c)** Sensitivity to face identity with and without adaptation to a non-opposite face. The proportion of correct responses at each identity level has been averaged across the four identity trajectories, fitting a four parameter logistic function to the data.

4.6.1 Methods

4.6.1.1 Materials: The target stimuli were generated in a similar manner to Experiment 5a, except that, following Rhodes and Jeffery (2006) the four 100% target faces (Dan, Jim, Rob, and Ted) were selected from a set of 20 faces to be somewhat dissimilar to each other. To do this, the Euclidean distance between every face was calculated selecting any four that were at least as far from each other as the average Euclidean distance between the twenty faces. Eight jittered versions of the 100% target face were also constructed in the same way as in Experiment 5a.

Following the construction of the 100% target faces, -80% opposite-adaptors, corresponding to the -80% anti-face used in Experiment 5a, were constructed for each of the targets. Additionally, opposite-face identity trajectories were constructed, extending between the 100% target identities and the -80% opposite faces in steps of 0.05 (4x37 faces). Non-opposite adaptors were then selected from a set of 30 faces (the 16 remaining faces from the original set of 20 plus an additional 14 faces) to be equally similar to the target as the -80% opposite face based on Euclidian distance. To do this, for each 100% target, a face from the set of 30 was selected that was the closest to the Euclidean distance between the 100% target and the corresponding opposite adaptor with the condition that no face could be selected twice. This method of generating the faces was based on the method used by Rhodes and Jeffery and appeared to result in the generation of very reasonable non-opposite adaptors that were both a very similar distance from the 100% target as the opposite adaptors but also a reasonable distance from the opposite adaptors. Finally, a non-opposite test trajectory, corresponding to the opposite identity trajectory, was constructed between the 100% target and the non-opposite adaptor (4x37 faces).

4.6.1.2 Procedure: For each parameter set, on each model, 100 simulations were run. On each simulation of a given model, the decision layer was first trained on the output of the face-space layer in response to the 8 jittered versions of the four 100% target identities. As with Experiment 5a, only the probability of the model making a correct response to a given target face was recorded.

On Opposite and Non-Opposite Baseline trials the probability of the model correctly identifying each of the 37 identity levels on each of the four opposite and non-opposite target trajectories was recorded (37 Levels x 4 target faces x 2 conditions). On Opposite Adaptor trials, the probability of the model correctly identifying each of the 37 identity levels on each of the four target trajectories following adaptation to the corresponding opposite-adaptor was recorded. On Non-Opposite Adaptor trials, the probability of correctly identifying each of the 37 identity levels on each of the four non-opposite trajectories following adaptation to each of the four non-opposite adaptors was recorded.

4.6.2 Experiment 6 Results

The probability of the model making a correct response at each identity level on each of the four trial types, Opposite Baseline, Non-Opposite Baseline, Opposite Adaptation, and Non-Opposite Adaptation, was averaged across the four target trajectories. A four parameter logistic function was fitted to the data for each condition for each simulation of the model. The effect of opposite and non-opposite adaptation was calculated relative to the baseline, taking the inflection point and subtracting it from the area under the baseline function.

Two qualitative criteria were considered; first, the value for the opposite adaptation should be greater than for the non-opposite adaptation, indicating that there was more adaptation in the Opposite than the Non-Opposite condition. Second, the opposite adaptation ought to result in a significant positive shift from the baseline, indicating that adaptation facilitated recognition in the Opposite condition. Note that, although in the study conducted by Rhodes and Jeffery (2006) there was also a significant facilitatory effect of adaptation in the Non-Opposite condition this was not considered essential, as there is no strong theoretical reason why a non-opposite face will necessarily facilitate recognition. Indeed, some non-opposite faces may be detrimental to the recognition of the target. To this end a two-tailed t-test was conducted on the data with the criterion for significance set to $p<0.01$. Parameter sets for which both criteria were fulfilled were scored as qualitatively correct, parameter sets for which one or both of the criteria were not fulfilled were scored as incorrect, and parameter sets for which both criteria were not significant were scored as null. For simplicity, the data were then collated across values of ϕ and θ (See Experiment 5a). As was found in Experiment 5a collating the results so that incorrect results were prioritised resulted in no change to the data for either the exemplar-based model or the two-pool norm-based model. Indeed, the qualitative maps for the two-pool and exemplar models indicate that across a very wide range of parameters these models made qualitatively correct predictions (Figure 4.11a & 4.13a). However, the qualitative map for the traditional norm-based model indicates that there was only a small range of parameters that gave qualitatively correct predictions (Figure 4.12a) and collating the results to prioritise incorrect results resulted in there being no qualitatively correct conditions at all.

An approximate quantitative best-fit was also calculated, fitting a four-parameter logistic function to the average data across participants on each parameter set, and calculating the SSE between these points and the corresponding points in the data taken from Rhodes & Jeffery (2006). Thus, the data presented in Figures 4.11b, 4.12b, & 4.13b correspond to the best fits obtained from each model. Note that, only the qualitative map corresponding to the best-fit value of α (which controls the degree of adaptation) is shown, though there was very little difference between the qualitative maps for a given model. An examination of the quantitative fits revealed that in line with the results found by Rhodes and Jeffery, adaptation to opposite and non-opposite adaptors facilitated identification of the target in all three models (indicated by a shift in the logistic function to the left of baseline). Moreover, it appears that adaptation to a non-opposite adaptor facilitated target identification less than adaptation to an opposite adaptor.

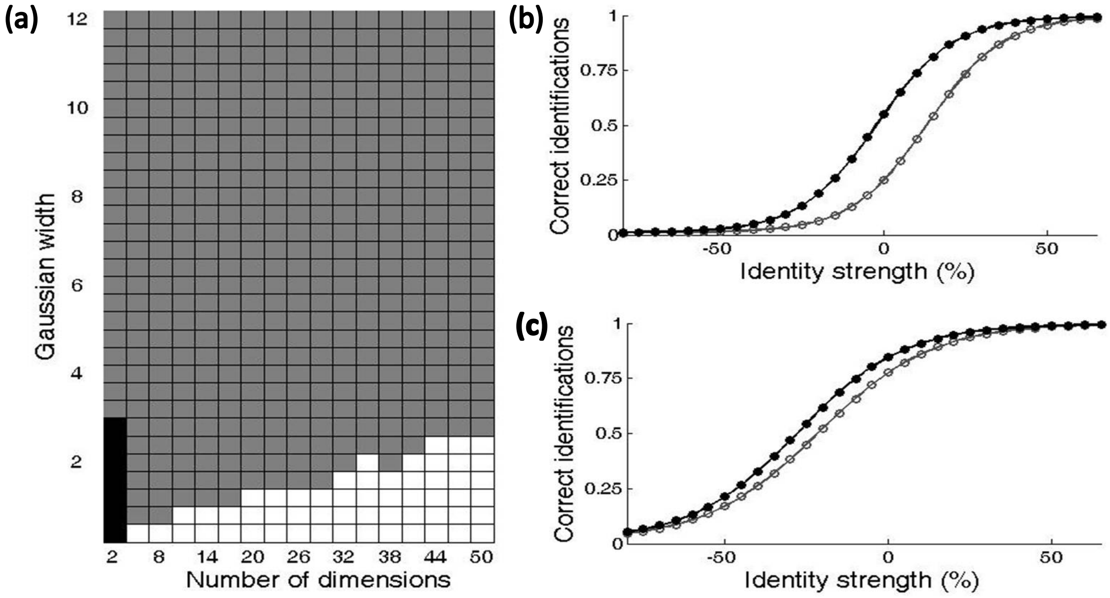


Figure 4.11. Simulation results from the exemplar-based model. **(a)** Map of the parameter space explored in Experiment 6 showing the qualitative fit with the findings reported by Rhodes and Jeffery (2006). Grey sqares (■) represent parameter values on which the model made qualitatively accurate predictions, white squares (□) represent parameter values that produced no signifcant adaptation (null), and black squares (■) represent parameter values in which the model made predictions that were qualitatively inconsistent with the findings of Rhodes and Jeffery. **(b)** Quantatative best-fit to the data from Rhodes and Jeffery opposite condition. Two conditions are shown, baseline responses (○), responses following adaptation to an opposite anti-face (•). **(c)** Quantatative best-fit to the data from Rhodes and Jeffery non-opposite condition. Two conditions are shown, baseline responses (○), responses following adaptation to a non-opposite adaptor (•). The best-fit parameter values were found for $n = 50$, $\eta = 5.6$, $a = 0.6$, and $\phi = 4$.

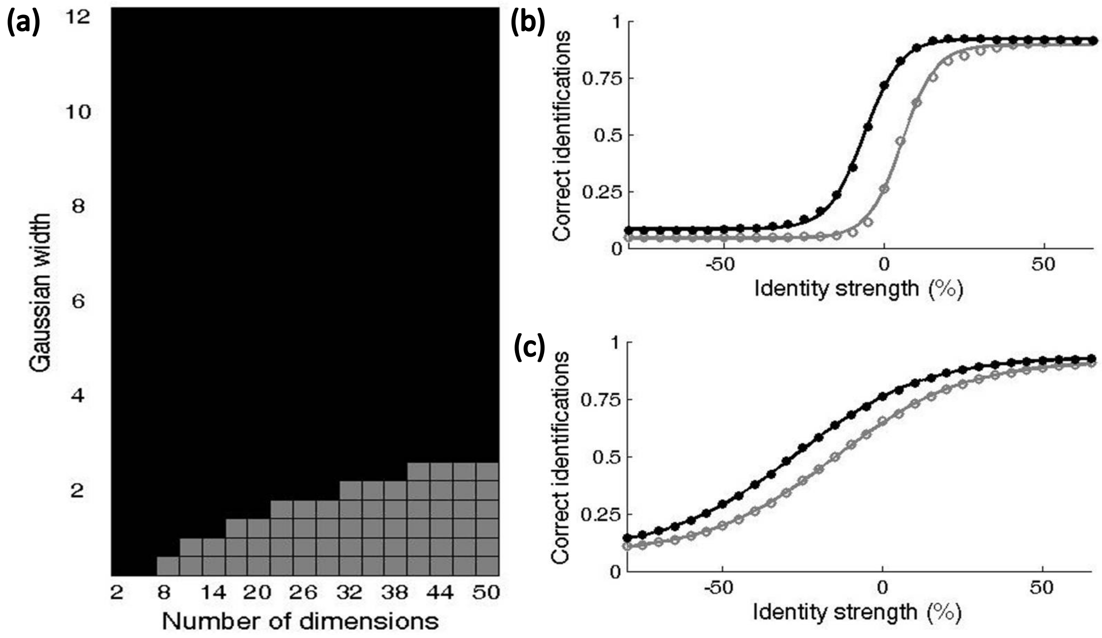


Figure 4.12. Simulation results from the traditional norm-based model (see Figure 4.11 legend for details). The best-fit parameter values were found for n = 50, v = 0.8, α = 0.8, ϕ = 2, and θ = *0.75*.

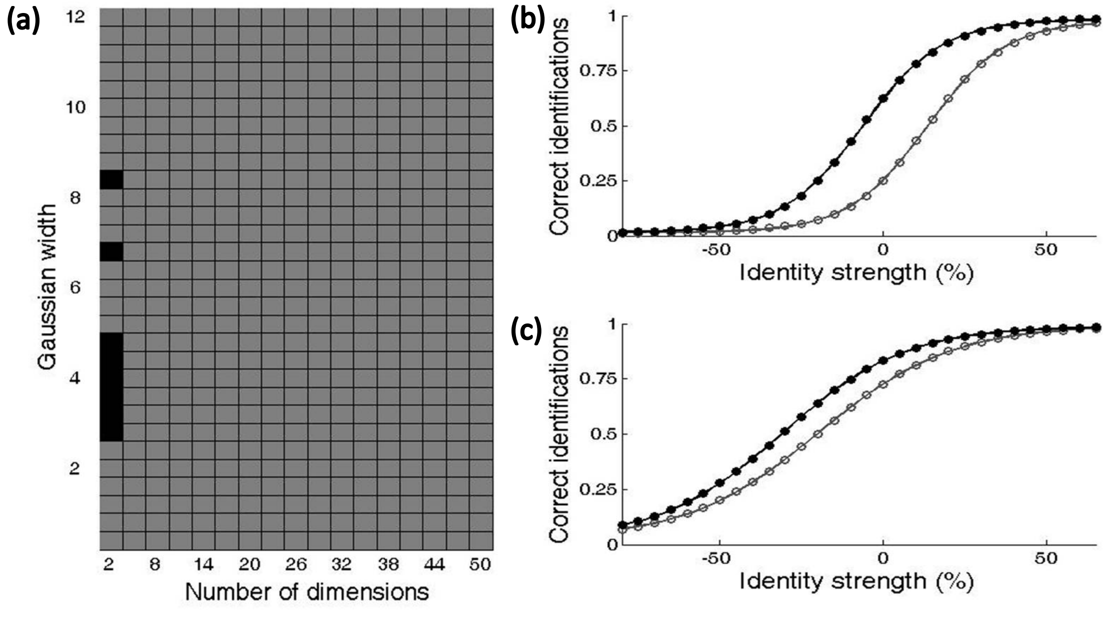


Figure 4.13*:* Simulation results from the two-pool norm-based model (see Figure 4.11 legend for details). The best-fit parameter values were found for n = 50, η = 10.4, α = 0.4, and ϕ = 4.

4.6.3 Experiment 6 Discussion

In Experiment 6, the three versions of face space, an exemplar-based face-space, a traditional norm-based face-space, and a two-pool norm-based face-space, were instantiated in order to investigate their qualitative predictions on the adaptation paradigm used by Rhodes and Jeffery (2006). Both the exemplar-based model and the two-pool norm-based model gave good qualitative fits across a broad range of parameters, indicating that adaptation to an opposite adaptor facilitated recognition of the target identity, and importantly, adaptation to opposite adaptors facilitated recognition more than adaptation to non-opposite adaptors. Indeed, in line with the data of Rhodes and Jeffery, the best-fit figures indicate that adaptation to both opposite and non-opposite adaptors facilitated recognition. However, this was not specifically tested for in the qualitative maps, as there was no strong theoretical reason to suspect that non-opposite adaptation would necessarily facilitate recognition. In contrast, the traditional norm-based model did not make qualitatively accurate predictions across a broad range of parameters. In fact, when the method of compiling the data across values of ϕ and θ θ was made more stringent, requiring that there were no qualitatively incorrect results for a given number of dimensions and Gaussian width, it was found that there were no qualitatively correct parameter values. Hence, the results reported in Experiment 6 would seem to provide evidence against this specific version of the norm-based model.

These results are important, particularly with respect to the exemplar-based model. The finding that adaptation to an opposite face facilitates the recognition of the 100% target identity more than adaptation to a non-opposite face is consistent with the idea that aftereffects are centred on the average face rather than on the adaptor. Thus, while the results are in keeping with the findings reported in Experiment 5c they extend the findings, suggesting that the exemplar-based model predicts that adaptation will be centred on the average face over a broad range of parameters. Indeed, other than in the two-dimensional space, all values of Gaussian width that resulted in any significant adaptation aftereffects also resulted in there being more adaptation in the opposite than the non-opposite condition. Thus, it would appear, given that the two-pool model also makes qualitatively consistent predictions, that the study conducted by Rhodes and Jeffery (2006) is insufficient to differentiate exemplar-based and two-pool norm-based models.

One aspect of the study that was not modelled here is that there was substantial learning in the non-opposite condition. To expand on this point, the paradigm used by Rhodes and Jeffery actually involved conducting baseline trials as one block, followed by adaptation trials as a second block and then a final block of baseline trials. Thus, the learning observed in their study appears to have taken place during the adaptation trials. Clearly as there is no allowance for learning in the model presented, it is difficult to speculate about what may have caused this learning. However, an important point to note is that while identification accuracy on the opposite trajectory is necessarily fixed at 25% at the 0% face, since all four trajectories pass through this point, identification at the equivalent point on the non-opposite trajectory is not fixed. This is because the non-opposite trajectories do not converge on any one point in face-space, though it is possible that they will cross. Indeed, while accuracy at the 0% identity level on the opposite trajectory was around 25%, accuracy at the equivalent point on the non-opposite trajectory was around 60%. As a result, while participants could not improve their accuracy at identifying the 0% identity level on the opposite trajectory they could have improved at identifying the 0% identity level on the non-opposite trajectory.

4.7 Experiment 7: Modelling Robbins et al. (2007)

In Experiment 7 a version of the paradigm used by Robbins et al. (2007) was instantiated in each of the three models. This paradigm provides a further test of whether adaptation is adaptor centred (as is typically assumed to be predicted by exemplar-based models) or centred on the average face (as is typically assumed to be predicted by norm-based models). As has already been discussed, one way to differentiate exemplar-based coding from two-pool coding is to take an adaptor at some point on a given dimension, say a face with eyes raised up by 20 pixels from the average, and examine its effect on the perception of targets that are both more extreme, say a face with eyes raised up 30 pixels from the average, and less extreme, say a face with eyes raised up 10 pixels from the average. As aftereffects are thought to be adaptor centred in an exemplar-based model the expected finding is that, following adaptation, the more extreme face will appear more extreme still, whereas, the less extreme face will appear less extreme still. In contrast, adaptation in a two-pool model predicts that adaptation is centred on the average. Thus, assuming that the targets are both on the same side of the average as the adaptor, (as they are in this example) both targets will be shifted in the same direction, that is to say, the direction of the average face, making them both appear less extreme.

To examine this, Robbins et al. created four faces, Bill, Sam, John, Fred. Each face was constructed as follows: the unique internal features were taken from the four faces and placed within the external features of a fifth face to create the four target faces. The faces were then distorted by moving both eyes up as far as the hairline and down as far as the bottom of the nose, creating faces with various levels of distortion between these extreme points. Next participants were taught the four target faces before being asked to rate how much like the original target each of the distorted versions of a given target were. Thus, the task was not a forced choice task; rather it simply involved rating how dissimilar, on a scale of 1 to 9, each of the distorted versions of Bill, Sam, John, and Fred were with respect to the original undistorted targets. In the adaptation conditions, the participants were adapted to a face at an intermediate level of distortion (e.g. either eyes slightly lower or eyes slightly higher than the original target) and then asked to rate faces at different levels of distortion in the same way as before. Note that, the adaptor face was not necessarily the same as the subsequently presented target. The important question being investigated here was whether aftereffects would be centred on the adaptor, with perception biased away from this location in both directions, or whether adaptation was centred on the average face, with perception biased towards the opposite distortion to the adaptor. The first of these possibilities would be in line with the predictions typically attributed to a two-pool model, whereas, the second of these predictions would be in line with the prediction typically attributed to a multi-channel model (for full predictions see Figure 4.14). The results suggested that adaptation to the intermediate face made faces that were more extreme appear closer to the original target, providing apparent support for the two-pool model.

Importantly, as this paradigm did not involve a forced choice decision it is possible to measure the effects of adaptation directly in the face-space layer, as in Experiment 5c, allowing a direct test of the predictions of each of the three models. Thus, as with Experiment 5c, the perceived location of each face before and after adaptation was measured using the centre of gravity of activation, extending the results of Experiment 5c over a broad range of parameters. Based on the other results reported in this chapter, and contrary to the theoretical analysis of Robbins et al (2007), it is expected that both the exemplar-based model and the two-pool norm-based model will make predictions that are qualitatively consistent with those reported by Robbins et al. The predictions for the traditional norm-based model are less clear.

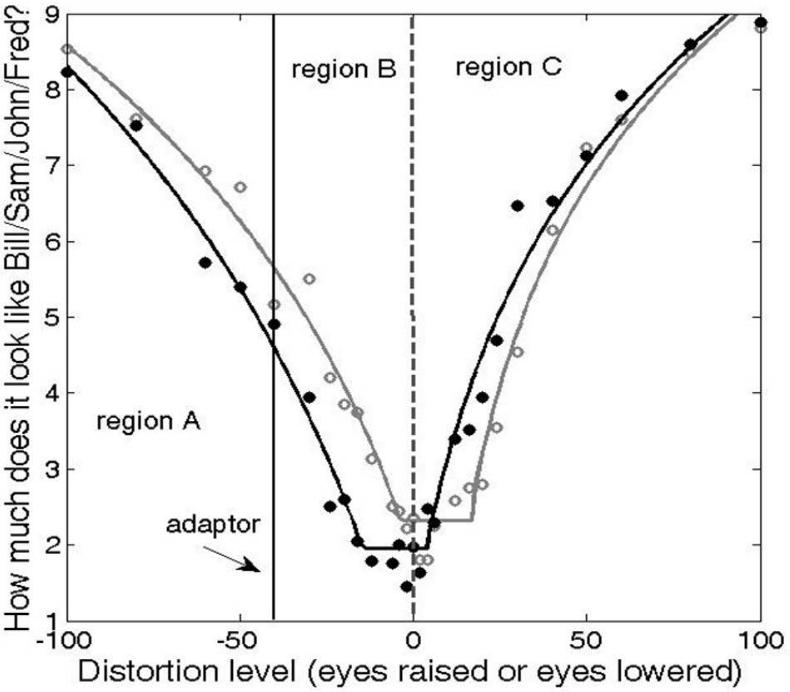


Figure 4.14. Dissimilarity ratings, 1=most like, 9= least like, versus distortion level, data taken from Robbins et al. (2007). Two conditions are shown, baseline responses (○), and responses following adaptation to a face at the -40% distortion level (●), equivalent to a face with the eyes slightly lowered towards the nose. **Region A:** distortion levels that are more extreme than the adaptor. **Region B:** distortion levels between the adaptor and the (0%) undistorted face. **Region C:** distortion levels on the opposite side of the original (0%) face. The predictions, based on a two-pool model are as follows: Region A faces will become more like the original target. Region B faces will also become more like the target but as this region is a transition region some may actually be adapted beyond the original target, that is to say their eyes may now appear higher than those of the target. Region C faces will become less like the original target. The expected predictions of an exemplar-based model are potentially the same except that faces in Region A would be expected to actually become less like the original target (Robbins et al.).

4.7.1 Methods

4.7.1.1 Stimuli: The stimuli used in this Experiment were generated to be as similar to those used by Robbins et al. (2007) as possible. As described, the stimuli used in their study were created by taking the internal features from four faces and placing them into the external features of a fifth face. In addition, the distortion, eyes up or eyes down, did not change the external features, or indeed, the perception of a substantial number of internal features. Thus, while the faces created for each simulation of the model were generated in the same way as for previous studies, selecting their location on each dimension from a random normal distribution, there were some substantial alterations. First, following the creation of four target faces, 60% of the dimensions were selected to be 'distorted features'. That is to say, features that were affected by moving the eyes up or down, and the location of each face on each of these dimensions was initially set to be equivalent to the origin. Second, 20% of the dimensions were set to be 'common features', simulating the shared external features of the four faces, thus the location of each face on these dimensions was identical, generated by taking the dimensions from a 5th randomly generated face. Last, the remaining dimensions were set to be 'individual features', simulating the unique characteristics of each face, indeed, these last dimensions were the only dimensions left unchanged from the original four generated faces. Note that, although these values have been used to generate the following simulation data their values appear to make very little difference to the qualitative predictions of the three models (this is further discussed in section 4.7.3).

Generating the distortion trajectory, -100 to 100% was done by first setting the values of the 100% distortion level on each of the 'distorted feature' dimensions. To do this the maximum absolute value of any face in the face-space on each of the 'distorted feature' dimensions was assigned to each of the four faces reassigning the positive or negative value. The reason for finding the most extreme values in the face-space was that moving the eyes up as far as the hairline or as low as the bottom of the nose produces faces that are at the extremes of the experienced face population (see Robbins et al., 2007). Distortion trajectories were then created for each of the four faces morphing between the 100% distortion level, through the original target face (0% distortion level), to the -100% distortion level in steps of 5% (4x41 faces). Note that, the 0% target does not correspond to the 0% identity level used in previous studies because 40% of the dimensions remained unchanged.

4.7.1.2 Procedure: For each parameter set on each model 100 simulations were run. *Baseline* - On each simulation, the centre of gravity (Equation 7) was calculated based on the activation in the face-space layer in response to each of the 41 faces in the four distortion trajectories. The Euclidean distance was then calculated between each COG, for each distortion level of each face, and the 0% distortion level of the corresponding face. *Adaptation* - The face-space layer was then adapted to the -40% distortion level of one of the four faces and the Euclidean distance between the COG of each face and the 0% level recalculated as in the Baseline condition. This procedure was repeated until each of the -40% distortion levels from each of the four distortion trajectories had been used as an adaptor. The Euclidean distances, *d,* were then converted into dissimilarity ratings as follows:

$$Dissimilarity = 1 - \exp\left(-\gamma d^2\right) \tag{8}$$

Whereby, γ was a parameter between 1 and 60 (varied in steps of 3)

4.7.2 Experiment 7 Results

Dissimilarity ratings for the Baseline and Adaptation conditions were averaged across the four distortion trajectories. Following Robbins et al. (2007) the effect of adaptation was calculated separately for each region, A, B and C by conducting a t-test on the averaged dissimilarity ratings for Baseline and Adaptation, $p<0.01$, to investigate if adaptation had significantly increased or decreased dissimilarity.

Three qualitative criteria were considered. First, in Region A, the dissimilarity ratings for the Adaptation condition should be lower (more similar) than in the Baseline condition, indicating that adaptation caused faces in this region to move closer to the 0% target. Second, in Region B, the dissimilarity ratings for the Adaptation condition should also be lower (more similar) than in the Baseline condition, indicating that adaptation caused faces in this region to move closer to the 0% target. Third, in Region C, the dissimilarity ratings for the Adaptation condition should be higher (more dissimilar) than for the Baseline condition, indicating that adaptation caused faces in this region to move further from the 0% target. The data were then collated across values of ϕ and θ and γ in the same way as in Experiment 5a and Experiment 6. Once again, testing revealed that collating them to prioritise incorrect conditions made no difference to the results of all three models. The results were then made into qualitative maps of the parameter space (see Figures 4.15a, 4.16a, & 4.17a). Examination of the qualitative maps indicated that across a very wide range of parameters all three models made qualitatively accurate predictions.

An approximate quantitative best-fit was also calculated, calculating the SSE between these points and the corresponding points in the data taken from Robins et al. (2007). Thus, the data presented in Figures 4.15b, 4.16b, & 4.17b correspond to the best fits obtained from each model. Note that, only the qualitative map corresponding to the best-fit value of α (which controls the degree of adaptation) is shown, though there was very little difference between the qualitative maps for a given model. An examination of the quantitative fits revealed that in line with the results found by Robbins et al. (2007), adaptation to the -40% adaptor caused the faces in Regions A and B to be rated as less dissimilar to the target and the faces in Region C to be rated as more dissimilar than before adaptation.

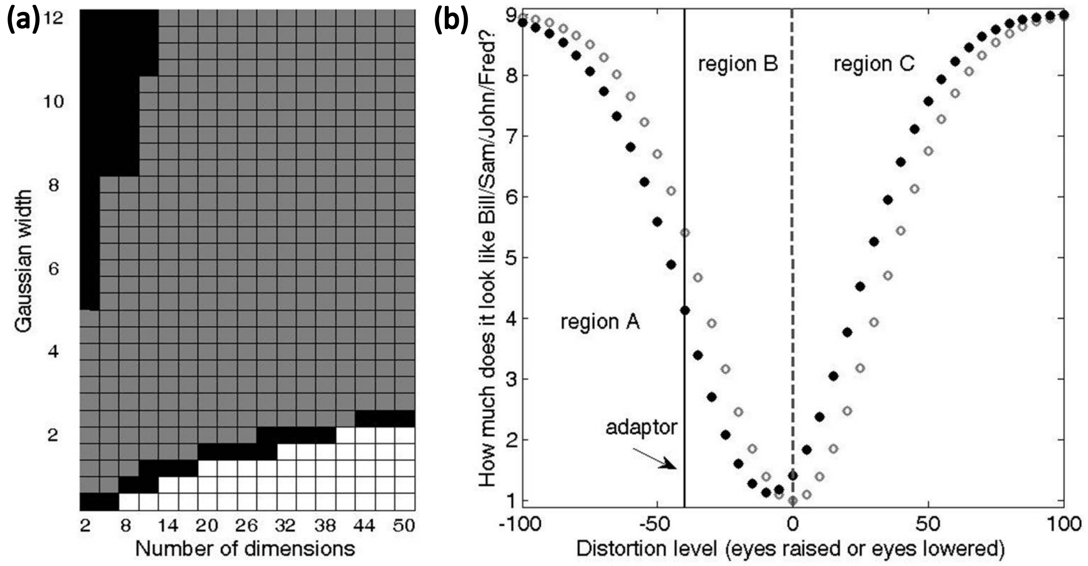


Figure 4.15: Simulation results from the exemplar-based model. **(a)** Map of the parameter space explored in Experiment 7a showing the qualitative fit with the findings reported by Robbins et al. (2007). Grey sqares (■) represent parameter values on which the model made qualitatively accurate predictions, white squares (□) represent parameter values that produced no signifcant adaptation (null), and black squares (■) represent parameter values in which the model made predictions that were qualitatively inconsistent with the findings of Robbins et al. **(b)** Quantatative best-fit to the data from Robbins et al. Dissimilarity rating, 1=most like, 9= least like, verses distortion level. Two conditions are shown baseline responses (○), and responses following adaptation to a face at the -40% distortion level (●). The best-fit parameter values were found for $n = 11$, $\eta = 2.8$, $\alpha = 0.8$, and $\Upsilon = 6$.

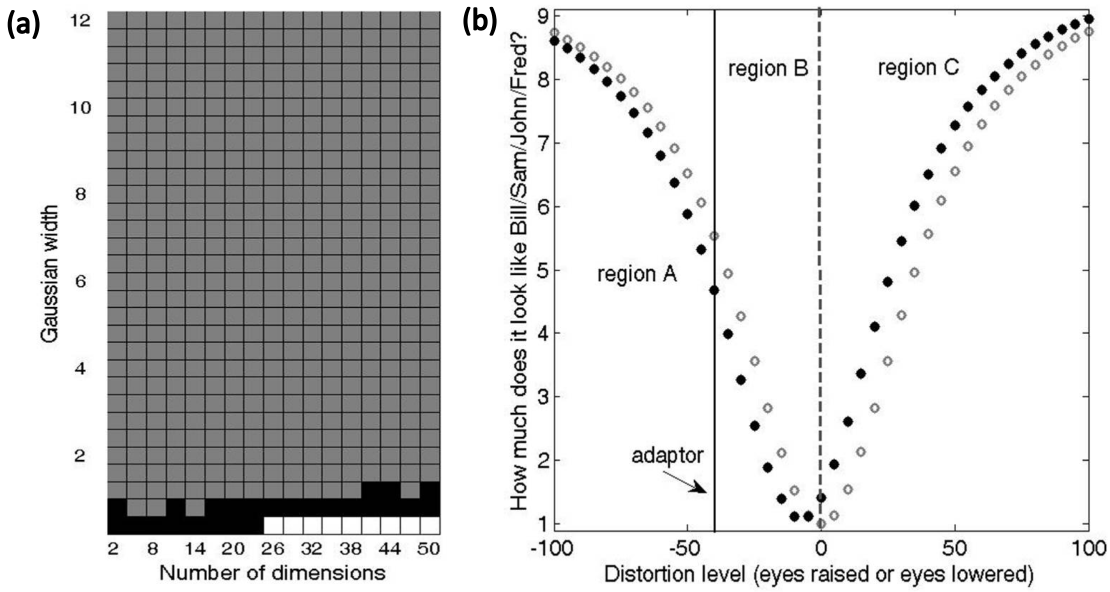


Figure 4.16. Simulation results from the two-pool norm-based model (See Figure 4.15 legend for details). The best-fit parameter values were found for n = 8, v = 2, α= 0.4, θ= *0.5,* and Υ= 2.

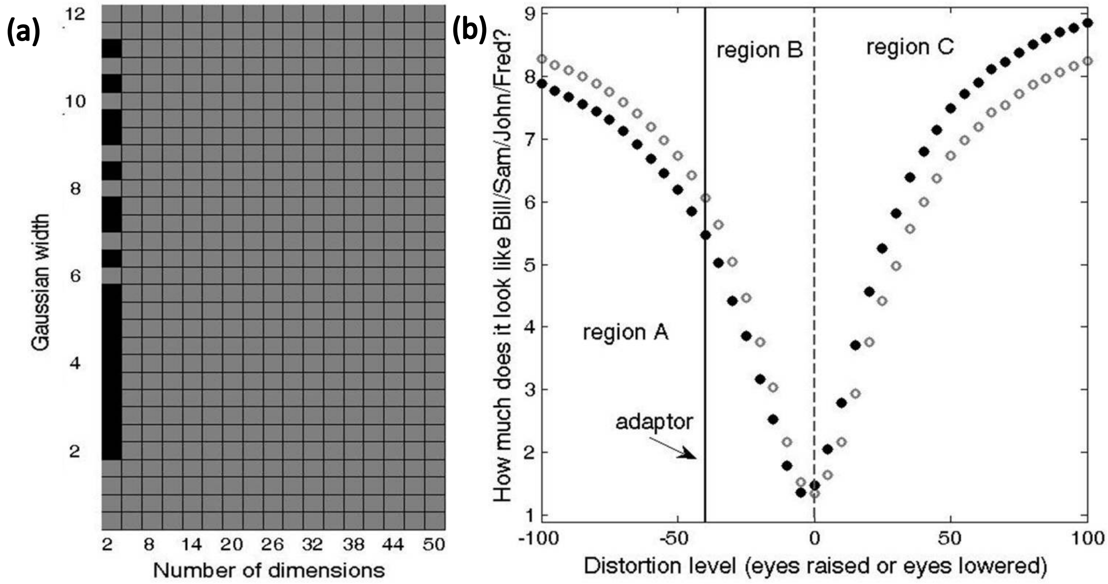


Figure 4.17. Simulation results from the traditional norm-based model (See Figure 4.15 legend for details). The best-fit parameter values were found for n = 5, η = 2, α = 0.4, and Υ= 14.

4.7.3 Experiment 7 Discussion

In Experiment 7, the three versions of face space were instantiated in order to investigate their qualitative predictions on the adaptation paradigm used by Robbins et al. (2007). All three models gave good qualitative fits across a broad range of parameters, indicating that adaptation to a -40% adaptor caused both more extreme and less extreme faces than the adaptor to appear more similar to the 0% distortion level. Furthermore, adaptors that were on the opposite side of the 0% distortion level to the adaptor became more dissimilar, indicating that they moved away from the 0%. Indeed, the best-fit figures indicate that all three models provided satisfactory best-fit data. One point that is worth noting is that the region of black, indicating an incorrect qualitative fit, on the upper left hand corner of the qualitative map for the exemplar-based model (Figure 4.15) is somewhat misleading. In fact, further testing of the model revealed that this region results from there being too much adaptation, rather than qualitatively wrong adaptation. Specifically, what appears to happen in this region is that faces in Region A become shifted so far in the direction of the 0% distortion level that they actually pass the 0% and become more dissimilar than they were before adaptation. This effect only occurs when there is a lot of adaptation (in this case a = 0.8). In the three qualitative maps corresponding to lower values of a (0.2, 0.4, and, 0.6) there was no region of black.

These results serve to confirm the rest of the findings reported in this chapter. They are particularly interesting with respect to the exemplar-based model as they extend the results reported in experiment 5c and Experiment 6, suggesting that over a very large range of parameters the exemplar-based model predicts that adaptation will be centred on the average face. The small region of black squares along the bottom of the qualitative map for the exemplar-based model do indeed represent conditions in which it makes qualitatively inaccurate predictions. However, two points worth noting are, first, the magnitude of adaptation in this region is very small, and second, the aftereffects only cover a very narrow range of the distortion trajectory. That is to say, while some of Region A and Region B are slightly affected by adaptation there are no significant differences in Region C. On this basis, it is possible to rule these parameter values out as they result in local adaptation aftereffects. Thus, adaptor centred adaptation appears to be the exception in an exemplar-based model and will only occur if exemplars are very narrowly tuned.

The construction of the stimuli in Experiment 7 was somewhat convoluted. This was in an attempt to accurately reflect the stimuli that were used by Robbins et al. (2007). In practice, the precise percentage of the dimensions that are assigned to be ‘distorted features’, ‘common features’ or ‘individual features’ seemed to be unimportant for the qualitative predictions. Moreover, a change to the percentage of dimensions assigned to each appears to affect all three models in a very similar way. Increasing the number of characteristic dimensions so that the distortion is applied to all dimensions results in a stronger adaptation aftereffect and reducing it results in a weaker adaptation aftereffect. If the number of dimensions is substantially reduced, the data also becomes noisier. The reason that reducing the characteristic dimensions increases the noise is that only these dimensions pass through the origin, the others lie some distance out. Thus, as aftereffects appear to be centred on the average, at least for the exemplar-based model and the two-pool norm-based model, only a very small component of the aftereffect goes in the direction of the distortion trajectory. Indeed, it was because of this quirk of the design used by Robbins et al. that it was felt necessary to allow these components to vary, ensuring that the models were able to tolerate the variation in the location of the target faces.

4.8 General Discussion

Taken together the simulations reported here represent a formal assessment of whether exemplar-based models are indeed falsified by existing data on face adaptation. Adaptation was implemented in three versions of face-space, an exemplar-based model, a traditional norm-based model, and a two-pool norm-based model. The predictions of these models were tested across the full range of experimental paradigms used to distinguish norm and exemplar accounts of face adaptation in the literature (e.g. Leopold & Bondar, 2005; Leopold et al., 2001; Rhodes & Jeffery, 2006; Robbins, et al., 2007; Susilo et al., 2010b). The main findings will now be summarised. First, the exemplar-based model and the two-pool norm based model made qualitatively accurate predictions on all tasks across a wide range of parameters. Second, the traditional norm-based model failed to predict the qualitative results reported by Rhodes and Jeffery across a wide range of parameters. Third, contrary to theoretical accounts of face adaptation the exemplar-based model did not predict that adaptation simply biased perception away from the adaptor location. Rather, for at least some parameters the exemplar-based model predicted that face aftereffects biased perception towards the opposite face, with the aftereffect increasing for adaptors further from the norm. Thus, it would seem that contrary to the theoretical claims made in the literature exemplar-based models might provide a neat account of face aftereffects.

Given the surprising nature of these results it is worth considering what might explain the results found here. Indeed, while it is not immediately obvious why the model predictions are not in line with theoretical accounts some clues might come from the findings reported in Experiment 5c and Experiment 7. First, only the broadness of tuning for each of the exemplar appears to be essential to the exemplar-based model making the prediction that adaptation will be average centred. If exemplars are very narrowly tuned then aftereffects begin to behave in a way that is more in line with previous theoretical accounts. That is to say, aftereffects become more centred on the adapting stimulus, biasing perception away from this point. Indeed, the region of black cells along the bottom of the qualitative map for Experiment 7 is the result of narrow tuning (Figure 4.15). However, it is worth noting that in these cases the adaptation is very local. That is to say, the aftereffects would only be noticeable for faces that were very similar to the adaptor. Thus, it would appear that for any parameters that are broad enough for adaptation to affect a broad region of the space the exemplar-based model would make qualitatively accurate predictions.

This observation raises the question of just how broad a region of the space it is desirable for a target face to activate. For example, the range of η (Gaussian width) used in the experiments reported in this chapter includes values much larger than the values used by Lewis (2004) in face-space-R. Indeed, as a result of the way that face-space-R decides if a target is familiar, that is to say, activation of the target minus the sum of the activation of all other exemplars, it would not be able to identify a familiar face at most of the values of η tested due to there being too much activation in competing exemplars. However, there would still be a small range of values for which the model predicts qualitatively correct adaptation aftereffects while still being consistent with face-space-R. Moreover, the assumptions that face-space-R makes, that is to say, that recognition is based on a winner-takes-all competition between exemplars might not be the only, or best, way of instantiating the exemplar-based model. Finally, it might be informative to compare the model predictions regarding exemplar tuning with physiological data on exemplar tuning in face selective cortical areas. Indeed, while the data is currently fairly sparse there have been both fMRI (Gilaie-Dotan & Malach, 2007; Jiang et al., 2006; Loffler et al., 2005) and single cell (Freiwald et al., 2009; Leopold et al., 2006) studies of face-selective cell tuning that would give a guide for model fitting. For example, one could compare the number of face units active in the model above a given threshold with the reported number of face-selective cells active in response to a given face (see Figure 4.18). Alternatively, one could look at release from adaptation in the model along various trajectory directions. The comparison of the model to neural data will be discussed more in Chapter 6. For now it is sufficient to conclude that the trade off between the size and direction of aftereffects in the model in combination with neural data would likely place a tight constraint on the range of parameters over which the model would fit the data.

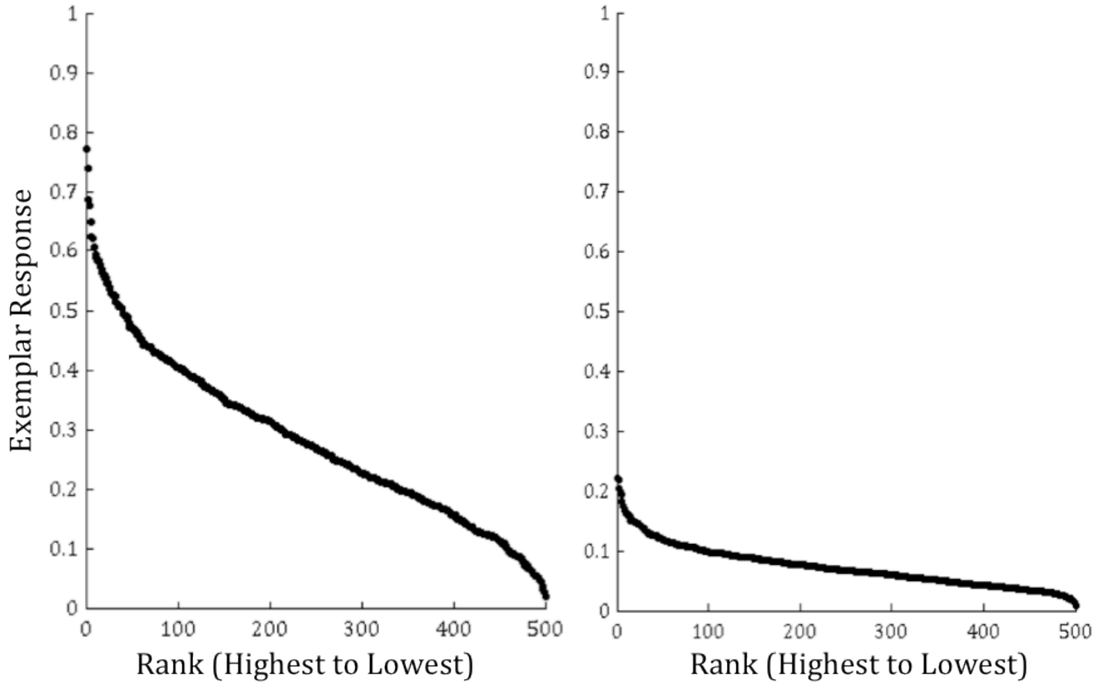


Figure 4.18. 500 face units ranked in order of their response to a 100% target identity versus the response of each face unit (exemplar response) response (between 0 and 1). **Left:** response of the 500 face units to a 100% target identity using the best-fit parameters found in Experiment 7. **Right:** Response of the 500 face units to a 100% target identity using the best-fit parameters for a 50 dimensional space in Experiment 7 (fits were nearly identical between the two cases). A response of 1 would result from the presentation of the units' preferred face. As discussed (see text) there is a broad range of parameters that produce qualitatively accurate fits to the behavioural data.

Another interesting finding that emerged from the simulations was that the traditional norm-based model made qualitatively incorrect predictions across a wide range of parameters in Experiment 6. In one sense this is surprising because theoretical interpretations of adaptation have generally assumed that aftereffects are consistent with norm-based coding. However, it is worth noting that adaptation was implemented in the traditional norm-based model by reducing the activity of the face units in proportion to their response to the adaptor face. In contrast, theoretical descriptions of norm-based models describe adaptation in terms of a change in the location of the norm relative to which all faces are encoded. Thus, these results do not directly contradict the theoretical descriptions; rather, they provide a demonstration of a possible implementation that is unable to predict the behavioural data. Indeed, shifting the location of the norm was beyond the scope of the implemented model, as it would require changing the input representation to the face-space layer.

To conclude, the modelling work presented in this chapter provides an interesting insight into possible models of face adaptation. The models implemented here represent the first attempt at exploring theoretical accounts of face adaptation in fully instantiated norm- and exemplar-based models. Surprisingly, it appears that contrary to previous claims, exemplar-based models can provide an accurate account of the empirical data. Indeed, this would seem to be an important point given that many authors consider the norm-versus-exemplar debate to be conclusively settled (e.g. Susilo et al., 2010a). Indeed, exemplar-based accounts have been entirely absent from recent reviews of face processing (e.g. Jeffery, & Rhodes 2011; Rhodes & Leopold 2011; Rhodes et al., 2005; Tsao & Freiwald, 2006). In addition, the current findings provide a confirmation of the theoretical claims about the two-pool model as well as some limitation on the range of norm-based models that would make predictions in line with the behavioural data.

Chapter 5

Adaptation in Novel Objects

Abstract

Chapter 5 reports three experiments that investigate whether average centred aftereffects, such as those found in studies of face adaptation, can be induced through brief exposure (over the course of a single experimental session) to a set of novel objects (Greebles). Experiment 8a investigated Greeble adaptation in a version of the face adaptation paradigm used by Leopold et al. (2001). Consistent with the results of face adaptation, adaptation to an anti-Greeble facilitated the identification of an opposite target-Greeble. Experiment 8b investigated if Greeble adaptation is centred on the average-Greeble, implementing a version of the face adaptation paradigm used by Leopold and Bondar (2005). In line with the results of face adaptation, the results of Experiment 8b indicated that adaptation to both a -50% and a -100% Greeble adaptor facilitated the recognition of the opposite target-Greeble more than adaptation to the 0% average-Greeble adaptor. Finally, Experiment 9 investigated Greeble adaptation in a version of the face adaptation paradigm used by Rhodes & Jeffery (2006). Contrary to findings in the face adaptation literature, adapting to an opposite Greeble-adaptor did not facilitate recognition of the target-Greeble more than adapting to a non-opposite Greeble-adaptor, indicating that Greeble adaptation did not bias perception towards a Greeble with opposite attributes to the average Greeble. Thus, it would seem that there might be some important differences between the aftereffects seen for novel objects and the aftereffects reported in the face adaptation literature.

5.1 Introduction

Average-centred[5] aftereffects are a hallmark of face adaptation (see Chapters 1 & 4), with adaptation to a given face biasing subsequent perception towards a face with opposite attributes to the adaptor (Rhodes & Leopold, 2011). Moreover, the magnitude of adaptation increases as the adaptor face is moved further from the average (Susilo et al., 2010a; Susilo et al., 2010b), with adaptation to an average face resulting in almost no bias in subsequent perception (e.g. Leopold & Bondar, 2005). However, some recent evidence has suggested that average-centred aftereffects, such as those that have been reported in studies of face adaptation, are not actually unique to faces (e.g. Dennett, Edwards, & McKone, 2009; & Belin, 2011; Susilo et al., 2010a).

In fact, average centred aftereffects have now been reported for inverted faces (Leopold et al., 2001; Robbins et al., 2007), horses (Dennett et al.), and even simple shape configurations, such as upright and inverted "T- shapes" (Susilo et al., 2010a). Indeed, using a version of the paradigm used by Rhodes & Jeffery (2006, see Chapter 4) to measure average-centred adaptation faces, a recent study even demonstrated average-centred adaptation in the perception of voice identity (Latinus & Belin, 2011). Importantly, these findings would seem to suggest that average-centred adaptive coding might be a fairly general principle in the high-level neural representation of category information (visual or otherwise). Furthermore, they raise the question of how much experience is

---

[5] In light of the findings reported in Chapter 4, in which both norm- and exemplar-based models were shown to account for many of the putatively 'norm-centered' aftereffects reported in the face adaptation literature, the term average-centered will be used throughout this chapter.

required with a given category in order to establish average-centred adaptive coding?

There is some indirect evidence that it might actually be possible to establish average-centred aftereffects in a given category of objects over the course of a single study session. For example, Susilo et al. (2010a) reported a study in which they investigated the transfer of aftereffects from upright and inverted T-shapes to upright and inverted faces. Their hypothesis was that, as upright faces are thought to activate face-specific neural populations, the transfer of upright face aftereffects to other objects ought to be minimal. In contrast, they suggested that as inverted faces are thought to be processed by face non-specific neural mechanisms (e.g. Aguirre, Singh, & D'Esposito, 1999) the transfer of inverted face aftereffects to other objects ought to be rather large. Indeed, this is what they found. By measuring the transfer of face-aftereffects from upright and inverted faces to upright and inverted T-shapes respectively they demonstrated that whereas 92% of the aftereffect for inverted-face-to-inverted-face also transferred to the inverted T-shapes, only 45% of the aftereffect for upright-face-to-upright-face transferred to the upright T-shapes. Thus, in line with their original predictions, they concluded that neural populations not specific to faces process inverted faces.

However, with respect to the question at hand it is relevant to note that, they also demonstrated that aftereffects for both inverted faces and inverted T-shapes were centred on the average stimulus. Specifically, they showed that the magnitude of the aftereffects produced by an inverted face or an inverted T-shape adaptor increased as the adapting stimulus was moved further from the average. Although they didn't comment on this finding it raises an interesting question; namely, if the neural populations that process inverted T-shapes and inverted faces are not specifically tuned for those stimuli, then why is it that they appear to encode information about the location of the average? Indeed, in the face adaptation literature it is assumed that aftereffects are centred on the average because this point has a ‘special’ status in the representation of faces, either representing a norm with respect to which faces are encoded, or the central tendency of the population of encoded exemplars.

One possible answer to this question is that the average-centred aftereffects that Susilo et al. (2010a) reported actually emerged over the course of the study session. Indeed, the T-shapes that they used in this study were based on the proportions of the internal features of an average face with the horizontal bar of the T-shape corresponding to the distance between the eyes on an average face and the vertical bar of the T-shape corresponding to the distance between the eyes and mouth on an average face. Thus, the fact that aftereffects were centred on the average T-shape suggests that participants must have learned something from exposure to the set of T-shapes over the course of the study. This would seem to be particularly true for the inverted T-shapes, as it seems unlikely that participants would have had extensive experience with such shapes, especially not experience that led them to have the same representation of the average as the one used in their study. Similarly, if, as they demonstrated, inverted faces are not encoded by face specific neural populations then it is unclear why the neural populations activated in response to inverted faces should encode the location of the average face. They key point here is that, setting aside the norm-versus-exemplar debate, it would seem that for aftereffects to be average-centred for a given set of stimuli, it is necessary that the neural populations encoding those stimuli are tuned to reflect the population statistics.

If it is the case that brief exposure to a stimulus set is sufficient to lead to norm-centred adaptive coding then, while it is normally assumed that face aftereffects are centred on the average face because this is the point that is central to our face-space representation, it is also possible that they are actually centred on the average because this point is central to the particular set of faces being tested. Some indirect evidence that this might be the case comes from the observation that, while face adaptation is always centred nearly perfectly on the average face, the average face appears to vary quite dramatically in its visual appearance across studies (Tsao & Freiwald, 2006, see Figure 5.1).

Thus, the studies reported in this chapter aimed to investigate whether average-centred aftereffects could be established in a completely novel set of stimuli, Greebles. It should be noted that while Greebles have a somewhat controversial status in the face recognition literature, frequently being used in studies arguing for an 'expertise' based account of numerous seemingly 'face-specific' phenomena (e.g. Gauthier & Tarr, 1997; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Tarr & Gauthier, 2000) this is not their intended use in the present study. Since it has already been established that average centred aftereffects are not face-specific (Dennett, Edwards, & McKone, 2009; Latinus & Belin, 2011; Susilo et al., 2010a) and, given that average-centred aftereffects have been shown for both horses and inverted T-shapes, it does not seem that being face-like is a pre-requisite. In this experiment the decision to use Greebles was based on the fact that they share a common configuration of parts such that it is possible to morph between one greeble and another in a smooth manner, allowing the various face adaptation paradigms to be implemented (see Figure 5.2). It is also important to note that while Greebles have been used to demonstrate ‘face-like’ effects, such effects only emerge over the course of many hours of training (e.g. Gauthier & Tarr, 1997). Thus, it is not something intrinsic to Greebles that leads them to be processed in a face-like manner, in fact, the results that have been found using Greeble stimuli have also been shown to emerge with expertise in cars (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 2000), Ziggerins[6] (Wong, Palmeri, & Gauthier, 2009) and even configurations on a chess board (Bilalic, Langner, Ulrich, & Grodd, 2011).

---

[6] Ziggerins are simple objects with similar basic constraints to Greebles. They look somewhat like cat scratching-poles.

Figure 5.1. Average face stimuli taken from three studies. **Left**: Average face taken from Leopold et al. (2001). **Middle**: Average face taken from Rhodes and Jeffery (2006). **Right**: Average face taken from Anderson and Wilson (2005). Despite all faces being putatively average there are clear differences between each of the faces used in these three different studies.

5.2 Experiment 8a: Anti-Greeble Aftereffects

In Experiment 8 a version of the face adaptation paradigm used by Leopold et al. (2001) was implemented using Greebles. As has been discussed in previous chapters (see Chapter 4), the face aftereffect reported by Leopold et al. is consistent with the idea of average-centred adaptation, with adaptation to an anti-face seemingly biasing perception towards the opposite identity. However, as has also been discussed, there is no way to rule out adaptor-centred aftereffects in which perception is simply biased away from the location of the anti-face (e.g. Rhodes et al. 2005). Thus, while Experiment 8a will not directly test whether Greeble aftereffects are centred on the average-Greeble it does provide a starting point for looking at adaptation in novel objects.

5.2.1 Method

5.2.1.1 Participants: The participants were a sample of 12 psychology undergraduates recruited from the Vanderbilt University participant panel and paid to take part in the study. All participants reported that they had normal or corrected to normal vision.

5.2.1.2 Materials: A set of 32 ¾-view images of symmetrical Greebles was compiled from a freely available database held at Carnegie Mellon University (http://www.cnbc.cmu.edu/ tarrlab/stimuli/novel-objects/index.html). Four, relatively dissimilar, Greebles were selected from the set with the remaining 28 morphed to construct an average-Greeble using Psychomorph (Tiddeman et al., 2001). The four remaining Greebles were then morphed 50% towards the average-Greeble (pre-testing indicated that a 50% morph Greeble was sufficient to get near ceiling-level identification). For simplicity, these four morphs will henceforth be referred to as the 'target identities', Aska, Teka, Luko, and Biki.

The construction of the experimental stimuli followed the method described by Leopold et al. (2001). Identity trajectories were constructed in 10% steps (identity levels) from each of the four target-identities (100% identity level) through the average-Greeble (0% identity level) to the -20% identity level on the other side of the average-Greeble (i.e. 4 identity trajectories x 13 identity levels). Thus, each identity trajectory corresponded to one of the four target identities. Four anti-Greebles were also constructed by morphing each of the four target identities through the average Greeble to create a -80% anti-Greeble on the other side of the average (anti-Aska, anti-Teka, anti-Luko, and anti-Biki, see Figure 5.2).

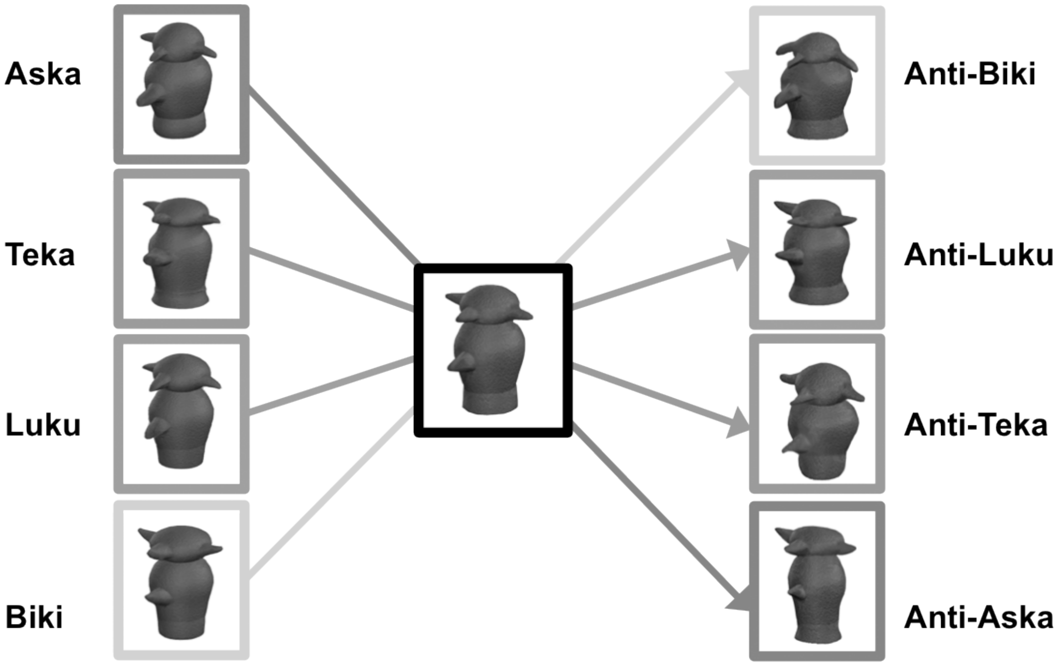


Figure 5.2. Stimuli used in Experiment 8a. **Left:** the four Greeble target identities. **Middle:** the average Greeble. **Right:** the anti-Greeble adaptors. The arrows represent the four identity trajectories passing through the average Greeble to the anti-Greeble on the other side.

5.2.1.3 Procedure: *Training* - The four target identities were presented individually, in a randomised order, along with the associated name, Aska, Teka, Luko, or Biki. Participants were allowed to study each image for as long as they wished before moving on to the next one by pressing the ‘a’, ‘t’, ‘l’ or ‘b’ key corresponding to the respective Greeble identity. This procedure was then repeated three more times, once more with the names present, and twice with the names absent. Following each trial, the participants received feedback as to whether they were correct or incorrect along with the correct name. Next, each target identity was presented on screen for 500ms without the associated name, followed by a blank screen that remained until the participant made a response. As before, the Greebles appeared in a random order and participants were provided with feedback. This phase lasted until the participant made 16 correct responses in a row.

Test Trials - There were four blocks of test trials, each block consisting of Baseline and Adaptation trials, with the order of the trials randomly intermixed within each block. The structure of Baseline trials and Adaptation trials was identical; a fixation-cross was presented for 250ms, followed by an adapting stimulus for 5000ms, a random noise mask for 250ms, a test stimulus for 500ms, and then a blank response screen. Baseline trials were designed to establish the threshold identity level that participants required to name the target identity in the absence of adaptation. Thus, on each baseline trial, the adapting stimulus was a blank white screen and the test stimulus was one of the 13 identity levels from one of the 4 identity trajectories. There were thus 208 Baseline trials in total (4 blocks x 4 identity trajectories x 13 identity levels). Following the presentation of the test stimulus participants were then prompted to identify target identity by pressing the appropriate key on the keyboard (‘a’, ‘t’, ‘l’ or ‘b’). Adaptation trials were identical except that the adapting stimulus on each trial was one of the four anti-Greebles, with each identity level from each of the four identity trajectories tested following adaptation to each of the four anti-Greebles. There were thus 832 Adaptation trials (4 blocks x 4 anti-Greebles x 4 identity trajectories x 13 identity levels).

5.2.2 Experiment 8a Results

Responses were scored as correct if they corresponded to the target identity from which a given identity level of test stimulus was constructed. For the purpose of analysis, the data were divided into three conditions: the Baseline condition, the Opposite Adaptation condition, and the Non-Opposite Adaptation condition. The Baseline condition comprised all of the Baseline trials. The Opposite Adaptation condition comprised those Adaptation trials in which the anti-greeble adaptor corresponded to the identity trajectory from which the test stimulus was constructed. That is to say, if the test stimulus was one of the identity levels on the Aska identity trajectory only trials in which the adapting stimulus was Anti-Aska would be included. Correspondingly, the Non-Opposite Adaptation condition comprised those Adaptation trials in which the anti-greeble adaptor did not correspond to the identity trajectory from which the test stimulus was constructed.

Following Leopold et al. (2001), a four parameter logistic function was fitted to the data for each condition (Baseline, Opposite Adaptation, Non-Opposite Adaptation) for each participant. The threshold identity level was taken at the inflection point of each logistic function and employed in subsequent analysis. Figure 5.3 shows logistic functions fitted to the mean data for each of the three conditions. A one-way repeated measures ANOVA revealed that there was a significant main effect of condition on identification threshold, *F*(2, 22)= 6.624, *p*= .006*.* Paired t-tests revealed that the difference between Opposite Adaptation threshold (mean=26.6%) and Baseline threshold (mean=35.7%) was significant, *t*(11)= 3.03, *p*= .01, as was the difference between Non-Opposite Adaptation (mean=38.7%) and Opposite Adaptation, *t*(11)= -2.75, *p*= .019. However, the difference between Non-Opposite Adaptation and Baseline was not significant, *t*(11)=1.15, *p*= .271.

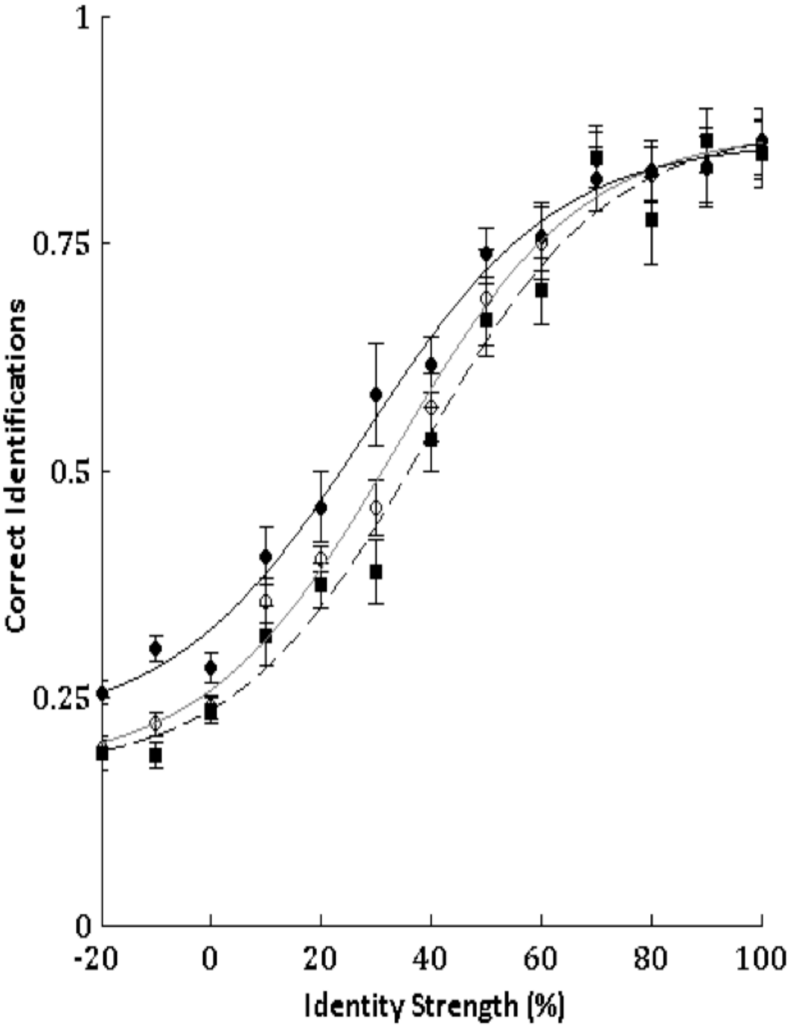


Figure 5.3*:* Sensitivity to identity in Experiment 8a. Three conditions are shown: Baseline (○), Opposite Adaptation (●), and Non-Opposite Adaptation (■). A logistic function has been fitted to the mean data points across participants in each condition with the inflection point used as a comparison across conditions (Error bars are SE for each identity level in each condition).

5.2.3 Experiment 8a Discussion

In Experiment 8a, a version of the face adaptation task used by Leopold et al. (2001) was implemented using Greeble stimuli. Participants were first trained to identify four target greebles over the course of a brief (10-20min) training session before undertaking the experimental session in which they had to identify morphed versions of the four targets both before and after adaptation to one of the four anti-Greeble adaptors. The results were generally consistent with the findings in the face adaptation literature, with adaptation to an anti-Greeble facilitating the recognition of the opposite target-Greeble. Indeed, in line with the results reported by Leopold et al. adaptation to a non-corresponding anti-Greeble (i.e. one which is not opposite) did not significantly facilitate the recognition of the target identity, if anything it may have slightly, though not significantly, hindered it.

These results provide a useful starting point for the investigation of adaptation in novel objects by demonstrating that qualitatively similar aftereffects can be found for Greebles as are found for faces. However, as has been discussed, the paradigm employed by Leopold et al. is insufficient to determine if adaptation is actually average-centred or if it merely resulted in a bias away from the adapting stimulus. Nevertheless, the correspondence found here between Greeble adaptation and face adaptation would suggest that some caution should be taken when interpreting the results of previous face adaptation studies. For example, several authors have used the correspondence between children's and adults' performance on this task as evidence that children have adult-like adaptive coding of faces (Nishimura et al., 2008; Nishimura et al., 2011; Pimperton et al., 2009). However, if it is possible to obtain similar aftereffects for a completely novel set of objects then it is unclear that a simple correspondence can tell us anything about face recognition. Indeed, using this task it might not even be possible to tell if someone has ever even encountered a face!

5.3 Experiment 8b: The Effect of Adaptor Position on the Magnitude of Greeble Aftereffects

In Experiment 8b, a version of the paradigm reported by Leopold and Bondar (2005) was implemented using Greebles. As has been discussed the paradigm reported by Leopold and Bondar is thought to provide a strong test of whether aftereffects are average-centred by measuring the magnitude of the aftereffect produced by adaptors which are at different distances from the norm (see Chapter 4). Specifically with respect to the present study, three levels of Greeble adaptor were introduced in place of the -80% anti-Greeble used in Experiment 8a, a 0% (average) Greeble-adaptor, a -50% Greeble-adaptor, and a -100% Greeble adaptor. If Greeble adaptation is centred on the average Greeble, as face adaptation is centred on the average face then it is expected that the magnitude of adaptation will be greater for adaptors that are further from the average, that is to say for the -50% and -100% Greeble adaptors.

5.3.1 Method

5.3.1.1 Participants: The participants were a sample of 10 psychology undergraduates, recruited from the Vanderbilt University participant panel, who were paid to take part in the study. All participants reported that they had normal or corrected to normal vision.

5.3.1.2 Materials: The set of four target identities, Aska, Teka, Luko, and Biki, as well as the average Greeble, were the same as those used in Experiment 8a. Identity trajectories were constructed in steps of 12% from each of the four target identities (100% identity level) through the average Greeble (0% identity level) to the -20% identity level on the other side of the average (4 identity trajectories x 10 identity levels). In addition, two levels of opposite adaptor were constructed for each identity trajectory, a -50% Greeble-adaptor and a -100% Greeble-adaptor.

5.3.1.3 Procedure: *Training & Test Trials* - The training procedure was identical to that used in Experiment 8a; test trials were as follows: There were four blocks of Adaptation trials, with the order of the trials randomly intermixed within each block. The structure of the trials was identical to that in Experiment 8a; a fixation-cross was presented for 250ms, followed by an adapting stimulus for 5000ms, a random noise mask for 250ms, a test stimulus for 500ms, and then a blank response screen.

Average adaptation trials were similar to the Baseline trials in Experiment 8a except that the adapting stimulus was the average (0%) Greeble-adaptor and the test stimulus was one of the 13 identity levels from one of the 4 identity trajectories. Thus, there were 208 Average adaptation trials (4 blocks x 4 identity trajectories x 13 identity levels). Following the presentation of the test stimulus participants were then prompted to identify target identity by pressing the appropriate key on the keyboard (‘a’, ‘t’, ‘l’ or ‘b’). The -50% and -100% Greeble-adaptor trials were identical except that the adapting stimulus on each trial was one of two levels (-50% or 100%) of one of the four Greeble adaptors. Unlike in Experiment 8a only the faces on the identity trajectory corresponding to a particular anti-Greeble were tested. Thus, there were 416 Adaptation trials (4 blocks x 2 levels x 4 identity trajectories x 13 identity levels).

5.3.2 Experiment 8b Results

For the purpose of analysis, the data was divided into three conditions: 0% adaptor condition, -50% Greeble-adaptor condition, and the -100% Greeble-adaptor condition. Responses were scored as correct if they corresponded to the target identity from which a given identity level of test stimulus was constructed. Following Leopold and Bondar (2005), a four parameter logistic function was fitted to the data for each condition (0%, -50% and -100% Greeble-adaptor conditions) for each participant. The threshold identity level was taken at the inflection point of each psychometric function and used in the subsequent analysis. Figure 5.4 shows a logistic function fitted to the mean data for each of the three conditions.

A one-way repeated measures ANOVA revealed that there was a significant main effect of condition on identification threshold, $F(2, 18)= 8.348$, $p= .0.03$. Paired t-tests revealed that thresholds in both the -50% Greeble-adaptor condition (mean=26.0%) and the -100% Greeble-adaptor condition (mean=23.3%) were significantly lower than in the 0% adaptor condition (mean=35.9%), $t(9)= 2.326$, $p= .045$, and $t(9) = 5.653$, $p< .001$, respectively. However, thresholds in the -50% Greeble-adaptor condition and the -100%

Greeble-adaptor condition were not significantly different *t*(9)= .919, *p*= .382 from one another.

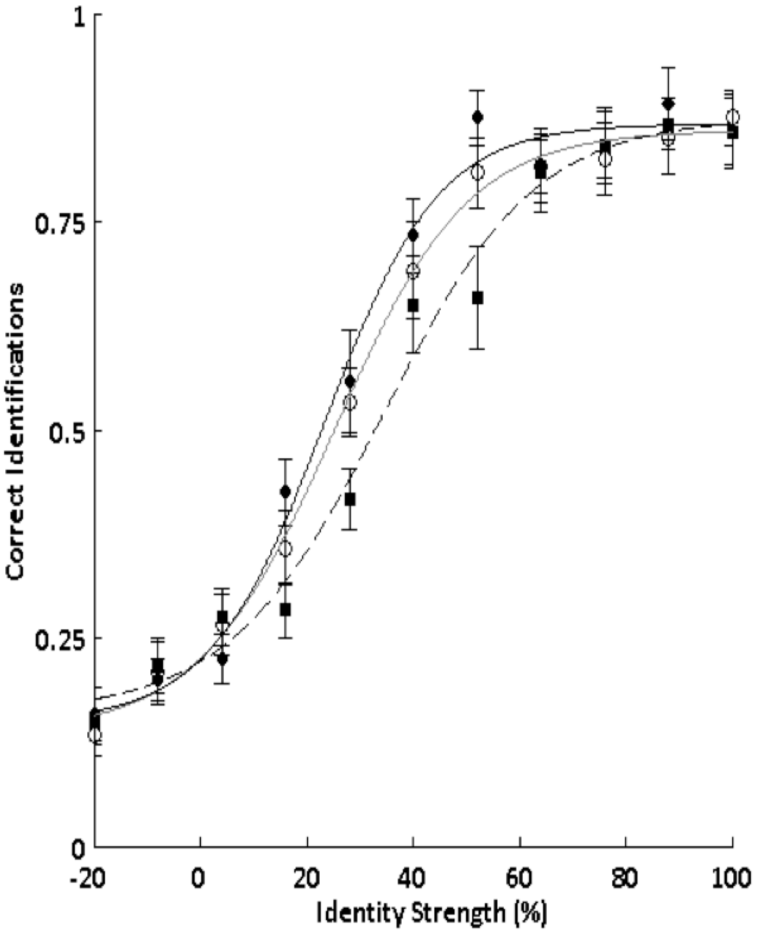


Figure 5.4: Sensitivity to identity in Experiment 8b. Three conditions are shown: -100% Greeble-adaptor condition (●), -50% Greeble-adaptor condition (○), and 0% Greeble-adaptor condition (■). A logistic function has been fitted to the mean data points across participants in each condition with the inflection point used as a comparison across conditions (Error bars are SE for each identity level in each condition).

5.3.3 Experiment 8b Discussion

In Experiment 8b, a version of the face adaptation task used by Leopold and Bondar (2005) was implemented using Greeble stimuli. Specifically, Experiment 8b investigated whether, in line with results in the face adaptation literature, the magnitude of Greeble aftereffects was larger following adaptation to Greebles that were further from the average Greeble. In line with this the identification thresholds were significantly lower for both the -50% Greeble-adaptor condition and the -100% Greeble-adaptor condition than they were for the 0% average Greeble-adaptor condition. However, the -50% and -100% Greeble-adaptor conditions were not significantly different from each other, perhaps indicating that the magnitude of the aftereffect levels off.

While these results do seem to be consistent with the findings in the face adaptation literature, indicating that Greeble adaptation is centred on the 0% average Greeble there is a significant methodological issue that would need to be addressed in any follow up studies. Specifically, given that no baseline (no adaptation) condition was included it is not actually possible to accurately measure the aftereffect size. What can be concluded is that both -50% and -100% Greeble-adaptor conditions led to lower identification thresholds than the 0% average-adaptor condition. With this in mind it is interesting to note that the threshold for the 0% adaptor condition was very similar to the baseline condition in Experiment 8a (35.9% and 35.7% respectively). However, given that the experimental design was not identical a direct comparison is not necessarily particularly informative.

Despite this limitation, the finding that adaptors further from the average facilitate recognition more than adaptors close to the average is consistent with average-centred adaptation. It is also inconsistent with a simple bias away from the adaptor, whereby, it is normally assumed that adaptation will decrease with an increase in the distance between the adapting stimulus and the target. Again, as with the results in Experiment 8a, these findings suggest that caution should be used when interpreting findings in the face adaptation literature. Indeed, it seems possible that with even brief exposure to a set of stimuli (possibly including faces) it might be possible to induce average-centred aftereffects.

5.4 Experiment 9: Measuring the Direction of Greeble Aftereffects

In Experiment 9, a version of the paradigm used by Rhodes and Jeffery (2006) was implemented using Greebles. In Rhodes and Jeffery's study it was found that opposite adaptor faces produced a greater shift in the threshold required for the identification of a target face than non-opposite adaptor faces, indicating that face adaptation biases perception in a direction that is opposite to the adaptor face (see Chapter 4). Following, Rhodes and Jeffery, the magnitude of the aftereffects produced by opposite and non-opposite Greeble-adaptors were measured. Opposite adaptors were constructed to be computationally opposite with respect to the average Greeble. Non-opposite adaptors were constructed to be equally far from the target-Greeble as the opposite adaptors but in a direction that was non-opposite with respect to the average Greeble. Identification thresholds were then measured along an opposite morph trajectory (between the opposite adaptor and the target Greeble) and a non-opposite morph trajectory (between the non-opposite adaptor and the target Greeble) both before and after adaptation to the respective opposite or non-opposite adaptor. Unlike the original study reported by Rhodes and Jeffery (2006) Experiment 9 was run with direction (opposite or non-opposite trajectory) as a between subjects factor.

5.4.1 Method

5.4.1.1 Participants: The participants were a sample of 32 psychology undergraduates, recruited from the Vanderbilt University participant panel, who were paid to take part in the study. All participants reported that they had normal or corrected to normal vision.

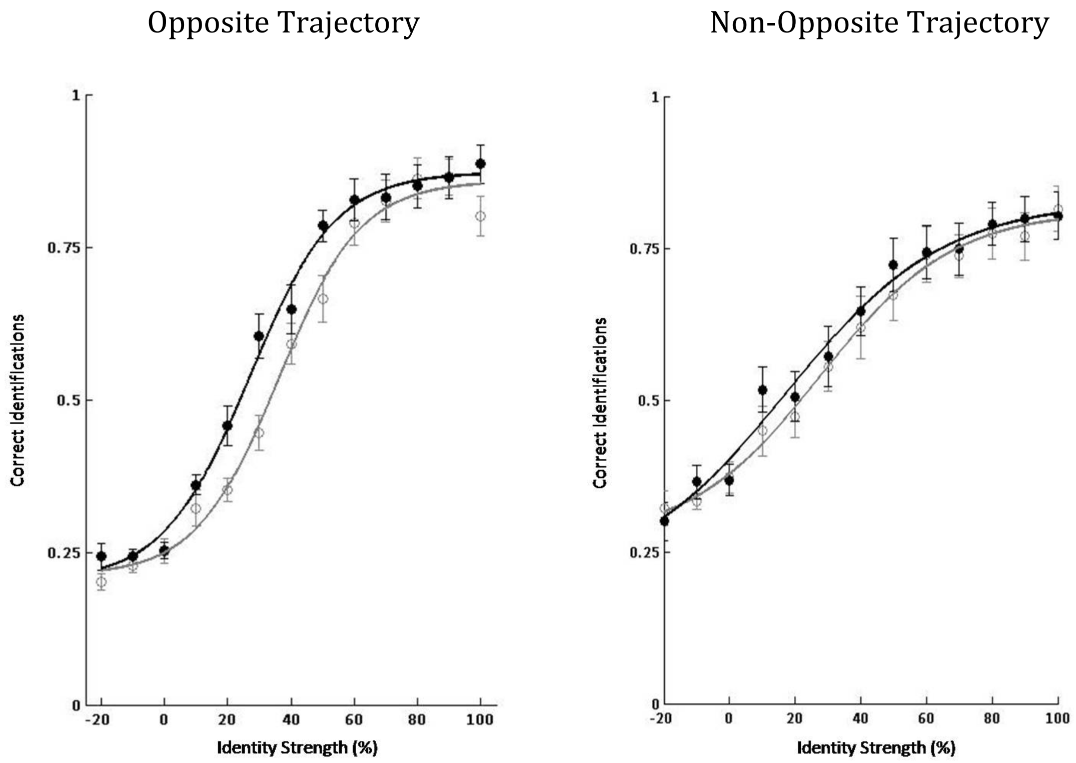
5.4.1.2 Materials: The set of four target identities, Aska, Teka, Luko, and Biki, as well as the average Greeble, were the same as those used in Experiment 8a. Opposite adaptors were created for each of the four target identities, constructing identity trajectories in 10% steps from each of the target identities (100% identity level) through the average Greeble (0% identity level) to the -20% identity level on the other side of the average (2 sets x 4 identity trajectories x 13 identity levels). In addition eight Opposite-adaptors were also constructed by morphing each of the target identities through the average Greeble to the -80% anti-Greeble on the other side of the average.

To construct Non-Opposite adaptors that were the same distance from the target faces as the Opposite adaptors the technique used to create the Oblique-Trajectory stimuli in Chapter 2 was applied. That is to say, the four targets were arbitrarily split into two pairs, Aska was paired with Teka and Luko was paired with Biki. Next the identity trajectory for one member of each pair, say Aska, was applied to the other member of the pair, in this case Teka. Thus, a morph trajectory was created extending from the 100% identity level of Teka with direction and distance specified by the Aska opposite morph trajectory. Correspondingly the identity trajectory for the other member of the pair, say Teka, was then applied to the first member, in this case Aska, in order to create the corresponding identity trajectory. Thus, the mean length of the opposite and non-opposite trajectories is the same.

5.4.1.3 Procedure: Participants were arbitrarily assigned to complete either an Opposite Adaptor version of the study or a Non-Opposite Adaptor version of the study, with 16 participants assigned to each condition. The participants in the Opposite Adaptor version of the study completed Adaptation trials with only Opposite adaptors and the participants in the Non-Opposite Adaptor version of the study completed Adaptation trials with only the Non-Opposite adaptors. The training and test trials were identical to those in Experiment 8a.

5.4.2 Experiment 9 Results

Responses were scored as correct if they corresponded to the target identity from which a given identity level of test stimulus was constructed. The data was divided into the four conditions, that is to say, the between subjects condition, direction (Opposite and Non-Opposite) and the within subjects condition, adaptor type (Baseline and Adaptation). Adaptation trials comprised only those trials in which the adaptor lay on the same trajectory as the corresponding target identity. Following Rhodes & Jeffery (2006), a four-parameter logistic function was fitted to the data for each participant in each condition at each level taking the threshold identity level at the inflection point of each psychometric function. Figure 5.5 shows logistic functions fitted to the mean data in each of the two conditions for both groups of participants. A repeated measures ANOVA with a within subjects factor, adaptor type (Baseline, Adaptation), and a between subjects factor, direction (Opposite, Non-Opposite), revealed that there was a significant main effect of adaptor type $F(1, 30)= 11.523$, $p= .002$, but no interaction between adaptor type and direction $F<1$. In addition, there was no between subjects main effect of direction $F(1, 30)= 2.101$, $p= .158$.



Opposite Trajectory

Non-Opposite Trajectory

Figure 5.5. Sensitivity to identity in Experiment 9. **Left:** sensitivity to identity in the Opposite trajectory condition before and after adaptation. **Right:** sensitivity to identity in the Non-Opposite trajectory condition before and after adaptation. Two conditions are shown in each plot: Adaptation (●) and Baseline (○). A logistic function has been fitted to the average data points across participants in each condition with the inflection point used as a comparison across conditions (Error bars are SE for each identity level in each condition).

5.4.3 Experiment 9 Discussion

In Experiment 9, a version of the paradigm used by Rhodes and Jeffery (2006) was implemented using Greebles. In line with the results from Experiment 8a and 8b, adaptation had a significant effect on threshold required for identification of a target face, facilitating recognition. However, in contrast to the results found by Rhodes and Jeffery, there was no indication that adaptation facilitated the identification of the target Greeble more in the Opposite condition than in the Non-Opposite condition. These results would seem to suggest that, in contrast to the findings in the face adaptation literature, adaptation in this study was not centred on the average Greeble.

One caveat to this conclusion is that there were some differences between the study reported here and the face adaptation study reported by Rhodes and Jeffery (2006). First, while it is not obvious that running the study between subjects should affect the results, it might be worth rerunning the study within subjects, particularly given that Experiment 8b seemed to support the possibility of average-centred adaptation in Greebles. A second difference is the way that the non-opposite trajectories were constructed. In the present study the non-opposite trajectory was created by morphing in an oblique direction, whereas, in Rhodes and Jeffery's study the non-opposite adapter was selected from a set of faces so as to be equally similar to the target as the opposite adaptor. Again, it is not obvious why this should affect the results, as the trajectories used here were still non-opposite but one possible difference is that the method used here actually resulted in the non-opposite greeble adaptors being vey distinctive with respect to the four target greebles, that is to say, the trajectories were very non-opposite. Indeed, if anything, this should have increased the possibility of finding a significant difference between the two trajectory directions, not decreased it. But without either using exactly same stimulus construction methods as those used by Rhodes and Jeffery, or using the stimulus construction method used here to conduct the study with face stimuli, this cannot be ruled out as a possible reason for the failure to replicate their results.

A final consideration is that different measures of average-centred adaptation might emerge at different stages of learning a particular stimulus set. Thus, it may be possible to see effects of adaptor distinctiveness on aftereffect magnitude for a set of stimuli but not find a difference between opposite and non-opposite adaptors. This would be an important finding but it would seem that it would take some further investigation before drawing such conclusions.

5.5 General Discussion

The three studies reported in this chapter set out to investigate whether it was possible to induce average-centred aftereffects in a set of novel objects over the course of a single experimental session. As was discussed in the introduction, it would be somewhat problematic if this was the case as it would suggest that such adaptation paradigms would reveal as much about the stimulus set as they would about previously established neural representations. Experiment 8a, while not directly addressing whether Greeble adaptation was average-centred, provided evidence of qualitatively similar adaptation aftereffects for Greebles as have been found for faces. Specifically, implementing a version of the paradigm used by Leopold et al. (2001) it was shown that adaptation to an anti-Greeble facilitated identification of the opposite target-Greeble. In addition, Experiment 8b provided some evidence that, in line with findings in the face adaptation literature, the magnitude of the aftereffect produced by a Greeble adaptor is related to its distance from the average Greeble. Finally, in Experiment 9, a version of the paradigm used by Rhodes and Jeffery (2006) was implemented using Greebles. In contrast to the previous two studies, the results of Experiment 9 were not in line with findings in the face literature, indicating that adaptation to an Opposite Greeble adaptor did not facilitate recognition of the target Greeble more than adaptation to a Non-Opposite greeble adaptor.

While these results do not convincingly demonstrate that the full gamut of ‘face-like’ aftereffects can be induced with a set of novel objects they do suggest that some caution should be taken when interpreting the results of certain face adaptation paradigms. For example, the results of the first two experiments (Experiments 8a & 8b) suggest that it is possible to obtain results that are qualitatively similar to those reported in the face adaptation literature, using a novel set of objects in two widely used paradigms (i.e. Leopold et al., 2001; Leopold & Bondar, 2005). In the first paradigm (Leopold et al), the basic finding is that perception can be biased away from the adaptor, though the direction is not specifically measured. This paradigm has been used in several studies comparing developmentally normal adults with prosopagnosics (Nishimura et al., 2010; Palermo et al., 2011) and children (Anzures et al., 2009; Nishimura et al., 2008; Pimperton et al., 2009). However, if the effect emerges spontaneously for a set of novel objects then caution should be taken when using such results to draw conclusions about face processing in these groups. Indeed, as discussed in the introduction to this chapter Susilo et al. (2010a) have demonstrated that some part of the inverted and upright face aftereffect can be attributed to face non-specific neural populations. The second paradigm used in this chapter (i.e. Leopold & Bondar, 2005) was essentially a demonstration that the magnitude of aftereffects can increase as the adapting stimulus is moved further from the average. Again, while the results reported here are not conclusive (see Experiment 8b discussion), they do suggest that caution should be taken when making claims based on this paradigm (e.g. Jeffery et al., 2010: Skinner & Benton, 2010; Susilo et al., 2010a; Susilo et al., 2010b).

The results reported in the third study (i.e. Rhodes & Jeffery, 2006) are interesting because unlike the first two studies the results were inconsistent with the findings in the face recognition literature. One possible explanation for this is that this paradigm is able to tap into something more fundamental about average-centred adaptation than the first two paradigms. Thus, it might be possible to investigate the development of stimulus specific expertise using adaptation aftereffects. For example, would 'face-like' aftereffects emerge after sufficient training? Presumably, one could also look at the degree of transfer of aftereffects to other similar shapes as a function of experience with a particular stimulus set. In addition one could use adaptation to look at the broadness of tuning of the neural populations responding to a given stimulus as a function of experience, or type of experience (e.g. categorisation vs. individuation).

In conclusion, the studies in this chapter show some important similarities between adaptation in a set of novel objects and adaptation in faces. However, these studies are only preliminary and some methodological issues mentioned in the discussion of Experiment 8b and Experiment 9 make it difficult to draw any firm conclusions. Nonetheless, these studies would seem to provide a ‘proof of concept’, suggesting that investigation of the development of ‘face-like’ adaptive coding in novel objects might be a productive direction for future research.

Chapter 6

Conclusions and Future Directions

6.1 Introduction

The purpose of this chapter is to provide an overview of all the experiments carried out over the course of this thesis, first recapitulating the central questions addressed in Chapter 1 and then summarising the experiments and findings reported in Chapters 2-5. In addition, where appropriate this chapter will provide a synthesis of the various lines of enquiry relating them to the broader theoretical framework around norm- and exemplar-based models of face recognition. Finally, some future research questions will be considered, ending with a concluding summary.

6.2 Summary of the Main Findings

Face-space models provide a useful framework for understanding human face recognition, unifying the effects of distinctiveness, caricature and, more recently, adaptation (e.g. Leopold et al., 2001; Lewis, 2004; Rhodes & Jeffery, 2006; Valentine, 1991; Valentine & Endo, 1992). Broadly speaking, face-space models fall into one of two categories. On the one hand, faces may be encoded with respect to a norm, representing how each face deviates from the central tendency. On the other hand, faces may be encoded with respect to their similarity to the population of familiar face exemplars as a whole. Unfortunately, empirically differentiating norm- and exemplar-based models has proved difficult, with both models accounting for many of the effects ubiquitous to face recognition.

It is generally thought that both norm- and exemplar-based models can explain why it is that typical faces are categorised, as faces, more rapidly than distinctive faces (Valentine and Bruce, 1986) and conversely, why it is that distinctive faces are recognised faster and remembered better than typical faces (Light, Kayra-Stuart, & Hollander, 1979; Valentine & Bruce, 1986). Moreover, both norm- and exemplar-based models offer neat accounts for the caricature effect. In the norm-based model caricature simply exaggerates precisely those features that are psychologically salient about a given face (e.g. Leopold et al., 2001; Rhodes & Jeffery, 2006; Rhodes et al., 1998; Tsao & Freiwald, 2006), whereas, in the exemplar-based model caricature moves the face to a region of lower exemplar density, with the effect that the target exemplar is activated proportionally more than it otherwise would be (Lewis & Johnston, 1998; Lewis 2004).

Recent research into face adaptation has led to a substantial literature in favour of some version of the norm-based model (e.g. Freiwald et al., 2009; Griffin, et al., 2011; Leopold & Bondar, 2005; Leopold et al., 2001; Rhodes & Leopold, 2011; Rhodes & Jaquet, 2011; Rhodes, et al., 2011; Rhodes & Jeffery, 2006; Rhodes et al., 2010; Rhodes et al., 2005; Robbins, et al., 2007; Susilo, et al. 2010b; Tsao & Freiwald, 2006). Face aftereffects appear to bias perception towards a face with ‘opposite’ attributes to the adaptor face (Robbins, et al.; Rhodes & Jeffery; Webster & MacLin, 1999), a finding that is consistent with the idea that the location of the norm, relative to which faces are represented, is moved. Furthermore, adaptation to an average face produces almost no aftereffect (Leopold and Bondar, 2005; Webster & MacLin, 1999) with aftereffect magnitude increasing based on the distance of the adaptor from the average face.

Importantly, exemplar-based models of face adaptation do not appear to make these predictions (e.g. Robbins et al., 2007; Rhodes & Jeffery).

One issue with the literature is the lack of a clear delineation between different versions of the norm-based model. For example, in some accounts of the norm-based model (the traditional norm-based model), it is suggested that face representations respond preferentially to a particular direction of deviation from the norm (e.g. Giese and Leopold, 2005; Leopold et al., 2001; Loffler et al., 2005; Blanz et al., 2000). Yet, not all accounts suggest that direction relative to the norm is important. Indeed, in some models the norm merely serves as a reference point for encoding the location of exemplars (e.g. Craw, 1995), while in others, such as the two-pool norm-based model, direction from the norm is implicit in the opponent coding of pools of face representations (Rhodes & Jeffery, 2006). Thus, the question of whether direction is important for face recognition was addressed in Chapters 2 and 3.

A second issue with the literature is that despite the strong claims made in support of norm-based models of face adaptation there have been no attempts to implement adaptation within norm and exemplar-based models. More generally, there have been very few attempts to implement a version of the norm-based model in any context (though see Giese & Leopold, 2005; Maytlis, 2011). This omission was addressed in Chapter 4, providing a ‘first pass’ at instantiating both norm- and exemplar-based accounts of face adaptation. Finally, though somewhat tangentially, Chapter 5 reported some experiments that looked at aftereffects in novel objects. While this approach did not directly address the central questions in this thesis it does, as will be discussed, have some interesting implications for studies of face adaptation.

6.2.1 Chapter 2: Discriminating Caricature and Oblique Face Pairs

As discussed in Chapter 1, traditional accounts of the norm-based model suggest that direction relative to the norm is explicitly encoded in the neural populations that represent faces. Importantly, this claim is not true of all versions of the norm-based model (e.g. Craw, 1995; Maytlis, 2011) or of exemplar-based models. In order to differentiate these accounts Chapter 2 reported two experiments that were designed to investigate whether upright (Experiment 1) and inverted (Experiment 2) familiar and unfamiliar faces are encoded with respect to their distance and direction of deviation from a norm (e.g. Giese & Leopold, 2005).

The approach taken in Chapter 2 was based on previous research into the effects of lateral and oblique caricature on face recognition (e.g. Lewis & Johnston, 1998; Rhodes et al., 1998). With the exception of a single unpublished study (Carey et al., 1992, cited by Carey, 1992) all previous studies have suggested that lateral or oblique caricatures are perceived as being less like the veridical face than standard caricatures and more like the veridical than anti-caricatures (e.g. Lewis & Johnston, 1998; Rhodes et al., 1998). Importantly, this pattern of results is widely accepted to be consistent with an account based on exemplar competition and inconsistent with an account based on direction relative to the norm (e.g. Lewis, 2004; Lewis & Johnston, 1998; Lewis & Johnston, 1999; Rhodes et al., 1998).

However, as mentioned earlier, despite the lack of empirical support from studies of lateral and oblique caricature, there still appears to be substantial support for a model in which direction relative to the norm is important. Thus, the studies presented in Chapter 2 were motivated by an attempt to control for some potentially confounding effects present in previous studies into lateral and oblique caricature. For example, in Chapter 2, the oblique caricatures were created so that they could reasonably be assumed to differ from each other in a way that represents the natural variation seen in a population of faces (Stevenage, 1997). In addition, the face stimuli created were also photorealistic, rather than line drawings as used by Rhodes et al. (1996).

In line with the idea that direction relative to the norm is important in the representation of faces, the results of Experiment 1 (upright familiar and unfamiliar faces) indicated that participants' discrimination thresholds were significantly lower for the Oblique-Trajectory pairs than for Caricature-Trajectory pairs. Thus, it appears that faces that lie on the same caricature, or identity, trajectory from the norm are perceived as more similar than faces that lie oblique to this trajectory. However, unlike Experiment 1, the analysis of the results from Experiment 2 (inverted familiar and unfamiliar faces) revealed that there was no significant main effect of trajectory direction for inverted faces, in addition, the interaction between direction and familiarity was also not significant. Unfortunately, the interpretation of these results is somewhat complicated by the fact that a combined analysis of Experiments 1 and 2 revealed a main effect of both trajectory direction and familiarity but no interaction with orientation. Nonetheless, it would seem that, contrary to much of the preceding literature (e.g. Lewis & Johnston, 1998; Rhodes et al., 1998; Rhodes et al., 2007), the studies presented in Chapter 2 provide some evidence in support of the traditional norm-based model, demonstrating that discriminations along the caricature/identity trajectory are easier than discriminations along other trajectories in face-space.

6.2.2 Chapter 3: Discrimination Across the Norm

In light of the results from Chapter 2, Chapter 3 reported two experiments that investigated a second, related, prediction of the traditional norm-based model. Namely, Chapter 3 investigated whether unfamiliar (Experiment 3) and familiar (Experiment 4) faces are perceived as opposite identities when they lie on different sides of the average face. The approach taken in Chapter 3 was similar to a number of previous studies (e.g. Blanz et al., 2000; Dakin & Omigie, 2009; Rhodes et al., 2007) measuring discriminability of face pairs that traversed a midpoint at different points along a given caricature trajectory. Importantly, a traditional norm-based model would suggest that faces on opposite sides of the norm ought to be perceived as opposites. In other words, a face-pair that traverses the norm is essentially an oblique-trajectory pair (e.g. Chapter 2) differentiated by the greatest possible angle relative to the norm.

Previous studies investigating discriminability across the norm have produced somewhat inconsistent results. For example, Rhodes et al. (2007) reported that face-pairs that traverse the average are actually perceived as more similar, whereas, Blanz et al. (2000) reported that they are perceived as less similar (in line with the traditional norm-based account). However, it should be noted that the results reported by Blanz et al. (2000) are of questionable reliability given that the step size in caricature across the average face was twice as large as the step size used elsewhere in the space (see Chapter 3 introduction). Nevertheless, consistent with Blanz et al., the results of Chapter 2 would seem to suggest that discriminability across the average face ought to be better than elsewhere along the caricature trajectory.

Although discrimination thresholds were measured at several points along the caricature trajectory, the most important prediction of the traditional norm-based model is that face-pairs that traverse the norm should be perceived as opposites. Consequently, a lower discrimination threshold for these face-pairs than for any other face-pairs would suggest that the angle between two faces is psychologically important. However, neither the results from Experiment 3 (unfamiliar faces), or Experiment 4 (familiar faces) showed any indication of better discrimination at any point along the caricature trajectory. Moreover, a combined analysis of Experiments 3 and 4, while revealing a main effect of familiarity, did not reveal a main effect of identity level, or an identity level by familiarity interaction. Thus, while the failure to find any significant effect of identity level on discriminability must be interpreted with some caution (as it is essentially a null result), it would seem reasonable to conclude, both from these findings and from findings reported in the literature (e.g. Dakin & Omigie, 2009; Rhodes et al., 2007; Susilo et al., 2010b) that there is little evidence that faces on opposite sides of the average are actually perceived as opposites.

6.2.3 Interim Summary: The Role of Direction in Face-Space

At first blush, the discrepancy between the results presented in Chapter 2 and those presented in Chapter 3 would seem problematic. Experiments 1 and 2 provided seemingly clear evidence of the importance of direction in participant similarity judgements, whereas, Experiments 3 and 4, in which the effect of direction on similarity should have been most pronounced, did not. However, it is notable that the results of Experiments 1 and 2 are quite different from the findings reported in previous studies. Indeed, previous research looking at lateral and oblique caricature has generally been consistent with an account based on exemplar density (e.g. Lewis & Johnston, 1998; Rhodes et al., 1998).

One way to picture the effect of exemplar density in an exemplar-based model (or in a norm-based model without direction coding) is in terms of a Voronoi diagram (Lewis & Johnston, 1999). Voronoi diagrams can be used to divide an n-dimensional space into different cells surrounding each face exemplar. In a two-dimensional space constructing a Voronoi diagram is fairly straightforward, involving drawing boundaries half way between a given exemplar and the next nearest exemplar in a given direction (see Figure 20). Each resulting cell of the Voronoi diagram represents the region in which the encompassed face is correctly recognised[7].

Voronoi diagrams provide a useful visual tool for understanding the effects of caricature on face-recognition. For example, as can be seen in Figure 6.1, exemplars do not generally sit centrally in each Voronoi cell. Rather, the boundaries of a given Voronoi cell are usually much further away from the exemplar in the direction directly away from the centre of the space than they are in the direction towards the centre of space. Indeed this observation corresponds neatly with the observation that caricatures are better recognised than anti-caricatures (e.g. Lewis & Johnston, 1999). Moreover, Voronoi diagrams can also be used to understand lateral, or oblique, caricatures given that the distance from the veridical representation to the edge of the Voronoi cell, that is to say the boundary perpendicular to the caricature trajectory, is generally shorter than the distance to the outer boundary (i.e. away from the centre of the

---

[7] Voronoi diagrams relate the effect of exemplar density to a classification decision meaning that any face would be classified as one of the selection of known familiar faces. While this is clearly untrue of face recognition and unsuitable for modeling similarity judgments or discrimination accuracy, Voronoi diagrams serve as a useful way to picture the behavior of an exemplar space.

space) and greater than the distance to the inner boundary (i.e. towards the centre of the space).

Considering Experiments 1 and 2 within the Voronoi framework it is clear that there may have been an important confound in the way that these studies were run. In previous studies (e.g. Lewis & Johnston, 1998; Rhodes et al., 1998) caricature trials were treated separately to anti-caricature trials, allowing a separate estimate of similarity for both types of distortion. In contrast, in Experiments 1 and 2, caricature and anti-caricature trials were combined, looking only at discriminability along the caricature trajectory versus discriminability along an oblique trajectory. Referring again to the Voronoi diagram in Figure 6.1, it is clear that this may have been an error. Specifically, one can think of the methodology used in Experiments 1 and 2 as a measurement of the length of a cell, measuring from the centre of the space directly outwards, and width of a cell, a measurement perpendicular to the length. Unfortunately, this method confounds the effect of caricature with the effect of anti-caricature and may have erroneously given the impression that there was an effect of direction on discriminability (given that cells are generally ‘longer’ than they are ‘wide’).

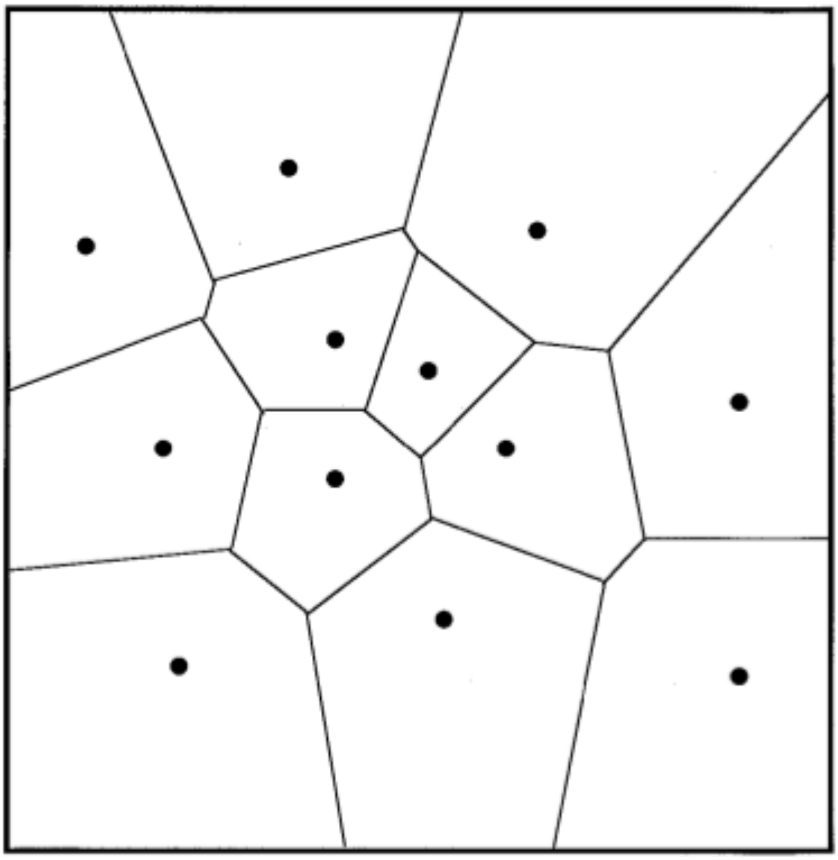


Figure 6.1. A Voronoi diagram constructed in a two-dimensional space with 12 exemplars (taken from Lewis & Johnston, 1999). The boundaries are constructed to be half way between each exemplar and its nearest neighbour with smaller cells near the centre of the space resulting from greater exemplar density. The off centeredness of each the cells has been used to understand the caricature effect.

However, although the above explanation provides a parsimonious account of the findings of Chapter 2, a complete explanation of the findings in the first two chapters remains elusive. Specifically, while the results would seem to rule out a traditional norm-based model it is not clear that they are in line with an exemplar-based model either. For example, inspecting the Voronoi diagram, it is apparent that cells near the centre of the face-space are generally smaller and less elongated than cells closer to the periphery. Following the logic used above to explain the results in Chapter 2, this would suggest that discrimination would be best in this central region, not equal, or even poorest as has been found by some studies (e.g. Dakin & Omigie, 2009; Rhodes et al., 2007; Susilo et al., 2010b). One solution to this problem is that increased similarity around the centre of face-space may emerge as a result of the highly overlapping exemplar representations. However, it would seem that without explicitly modelling the assumptions that underlie similarity ratings in an exemplar-based model it would be difficult to know how well it would predict this pattern of results. And the same could be said when relating the results to a traditional norm-based model, although it does seem that the predictions of the traditional norm-based model would be more straightforward, predicting that faces on opposite sides of the average should be perceived as opposite.

6.2.4 Chapter 4: Modelling Face Adaptation

Chapter 4 reported six studies that were designed to test the veracity of the theoretical claims made about face adaptation in norm- and exemplar-based face-space models. Face adaptation was implemented in three different versions of face-space: specifically, an exemplar-based face-space (e.g. Lewis, 2004), a traditional norm-based face-space (e.g. Giese & Leopold, 2006), and a two-pool norm-based face-space (e.g. Rhodes & Jeffery). The three different instantiations of face-space were then tested on a range of paradigms that have been purported to qualitatively distinguish norm- and exemplar-based models (i.e. Leopold et al., 2001; Leopold et al., 2005; Rhodes & Jeffery, 2006; Robbins et al., 2007; Susilo, et al., 2010a; Suzuki, 2005), assessing the qualitative predictions of each model across a wide range of parameters.

The first experiment, Experiment 5a, investigated the predictions of the three models in a version of the task used by Leopold et al. (2001). Leopold et al. demonstrated that adaptation biases perception towards a face with opposite attributes to the adaptor face. For example, recognition of the target face is facilitated by adaptation to the anti-face; moreover, adaptation to the anti-face facilitates recognition more than adaptation to an arbitrary non-opposite face. As with the original study, the aim of Experiment 5a was to compare the naming accuracy of the four target-faces at different identity levels both before and after adaptation to the corresponding anti-face. In line with the original findings reported by Leopold et al., all three models predicted that adaptation to an opposite adaptor would facilitate subsequent recognition across a broad range of parameter values. In addition, all three models predicted that a non-opposite adaptor would facilitate target-face recognition less than adaptation to an anti-face. Thus, from these results it would appear that the study reported by Leopold et al. is insufficient to differentiate the three models.

The fact that Experiment 5a did not reveal any qualitative differences between the three versions of face-space is not all that surprising. As a number of authors have pointed out, Leopold et al.’s (2001) results can be explained by both a bias in perception away from the adaptor, as is generally thought to characterise an exemplar-based model, or as a bias in perception in the opposite direction to the adaptor, as is generally thought to characterise a norm-based model (e.g. Rhodes & Jeffery, 2005). However, a simple extension to the study, reported by Leopold and Bondar (2005), is thought to be more diagnostic. In the extended version of the task, participants are adapted to an average face, rather than an anti-face. The logic follows that, if face adaptation produces a repulsive aftereffect then adaptation to an average face will bias perception away from the average, facilitating identification of the target, however, if face adaptation produces an opposite aftereffect then adaptation to an average face will not bias perception at all and target identification will be unchanged from baseline (e.g. Susilo et al., 2010b; Rhodes & Jeffery, 2006). Contrary to the above claims, an instantiation of this study carried out in Experiment 5b using the Experiment 5a best-fit parameters, revealed that all three face-space models predicted that adaptation to an average face would not facilitate identification of the target.

The results of Experiment 5b call into question the widely accepted claim that exemplar-based models predict that adaptation should be biased away from the location of the adaptor. One possible issue is that, these two studies only looked at the effect of adaptation on correct target identification. In order to get a clearer picture of the nature of aftereffects in an exemplar-based model it is also useful to look directly at exemplar activation. Thus, in Experiment 5c, the effect of face adaptation was directly computed from exemplar activation in the face-space layer of the model, determining the central tendency of activation produced by different target faces lying on a single arbitrarily selected trajectory through the average face. The central tendency was calculated for each target both before and after adaptation to an adaptor face, which lay at selected points along the same trajectory, plotting the difference. A comparison of the resultant plot with the adaptation tuning functions predicted for a norm- and exemplar-based model (e.g. Suzuki, 2005) clearly revealed that the exemplar-based model did not behave as was predicted in the literature. Indeed, it would appear from the results of Experiment 5c, that, much like a norm-based model, exemplar-based models predict that face adaptation produces a bias in perception towards a face with opposite attributes to the adaptor. Moreover, Experiment 5c also suggested that the magnitude of aftereffects in the exemplar-based model would increase as the adaptor is moved further from the norm.

In Experiment 6, a version of the paradigm used by Rhodes and Jeffery (2006) was instantiated in each of the three models. This paradigm, while similar to the one used by Leopold et al. (2001), provides an important extension of previously discussed studies, allowing the direction of the observed aftereffects to be inferred. As with Experiment 5a, the aim of Experiment 6 was to compare the naming accuracy of four target-faces at different identity levels both before and after adaptation to opposite (anti-face) and non-opposite adaptors. Importantly, unlike Experiment 5a the effect of adaptation was not only measured along the opposite trajectory - between the target and the opposite adaptor - but also along a non-opposite trajectory – between the target and the non-opposite adaptor. Thus, if face adaptation causes a bias in perception away from the location of the adaptor (as is claimed for an exemplar model) then both opposite and non-opposite adaptors should facilitate recognition of the target identity equally. But, if adaptation causes a bias in perception towards a face with opposite attributes to the adaptor (as is claimed for a norm-based model) then adaptation to an opposite adaptor should facilitate recognition more than adaptation to a non-opposite adaptor. Indeed, the latter prediction is consistent with the behavioural results reported by Rhodes & Jeffery, leading them to conclude that face-space is norm-based.

It is interesting then, that the simulations run in Experiment 6 demonstrated that both the exemplar-based model and the two-pool norm-based model predict that adaptation will be stronger along the opposite trajectory than along the non-opposite trajectory. Indeed, as with Experiment 5a, both of these models gave good qualitative fits to the behavioural data across a broad range of parameters. In contrast, the traditional norm-based model did not make qualitatively accurate predictions across a broad range of parameters. In fact, when the method of compiling the data across parameters was made more stringent, requiring that there were no qualitatively incorrect results for a given number of dimensions and Gaussian width, it was found that there were no parameter values that gave qualitatively correct results. Hence, the results reported in Experiment 6 would seem to rule out the version of the traditional norm-based model that was instantiated here.

In the final study, Experiment 7, a version of the paradigm used by Robbins et al. (2007) was instantiated in each of the three models. This paradigm provided a way of probing the precise direction of aftereffects in norm- and exemplar-based models, probing the activation in the face-space layer directly. Robbins et al. reasoned that if adaptation causes a bias in perception away from the adaptor location then the direction of an observed aftereffect will depend on the adaptor position relative to the target. Specifically, by creating a morph trajectory passing through the average between two extremes - eyes raised to hairline & eyes lowered to nostrils – it was possible to measure the effect of an intermediate level adaptor (e.g. eyes slightly raised) on both a more typical morph and a more distinctive morph on the trajectory. A bias in perception away from the adaptor would predict that the more typical morph would shift towards the opposite side (e.g. eyes lowered) and the distinctive morph would become more distinctive (e.g. eyes raised). Conversely, an opposite bias would predict that both the more typical morph and the more distinctive morph would be moved towards the opposite side of the average (e.g. eyes lowered).

As with Experiment 5c, the effect of adaptation in Experiment 7 was measured directly from the face-space layer of the model both before and after adaptation, using the central tendency of activation to determine the perceived location of each morph along the target trajectory and calculating the similarity to a reference face at the centre of the morph trajectory. Thus, experiment 7 provided a useful extension of the results of Experiment 5c, expanding the predictions to cover a wide range of different parameters and testing all three models. In line with the other findings reported in Chapter 4 but contrary to the claims of Robins et al. (2007), all three models, that is to say, both norm-based models and the exemplar-based model made predictions that were qualitatively consistent with those in the original study. Specifically, both more distinctive and more typical targets were biased towards the opposite side of the average in all three models.

To summarise, the results of Chapter 4 revealed that both the two-pool norm-based model and the exemplar-based model make predictions that are qualitatively consistent with the findings reported in the face adaptation literature. Moreover, these results were shown to be very general predictions of two-pool and exemplar-based models, with both models predicting the same qualitative pattern of results over a broad range of parameters. Additionally, it would seem possible to rule out the version of the traditional norm-based model implemented in Chapter 4, as it was unable to predict the qualitative pattern of results on the paradigm instantiated by Rhodes and Jeffery (2006).

These results are important because they challenge the currently dominant view that face adaptation is inconsistent with an exemplar-based model (e.g. Freiwald et al., 2009; Griffin, et al., 2011; Leopold & Bondar, 2005; Leopold et al., 2001; Rhodes & Leopold, 2011; Rhodes & Jaquet, 2011; Rhodes, et al., 2011; Rhodes & Jeffery, 2006; Rhodes et al., 2010; Rhodes et al., 2005; Robbins, et al., 2007; Susilo, et al. 2010b; Tsao & Freiwald, 2006). Indeed, as reviewed in Chapter 1, at present the results of such adaptation studies appears to be the only behavioural evidence that strongly favours a norm-based account of face recognition. If it is the case that exemplar-based models also predict the full range of face adaptation aftereffects then it would reopen the debate surrounding norm- and exemplar-based models.

6.2.5 Chapter 5: Adaptation in Novel Objects

Chapter 5 reported three experiments that investigated adaptation aftereffects in novel objects (Greebles). Specifically, the experiments reported in this chapter sought to determine whether the average-centred aftereffects seen in studies of face recognition might be induced with limited exposure to a novel set of objects. On the one hand, face aftereffects may, as is generally assumed, reflect statistical information represented in high-level visual areas, acquired through long-term experience with faces. On the other hand, it is possible that face aftereffects merely reflect statistical information acquired through short-term experience with a specific stimulus set.

Some evidence for the later hypothesis comes from the observation that, while average-face stimuli vary greatly in appearance across studies, face aftereffects always appear to be centred on the experimentally defined average face (Tsao & Freiwald, 2006). While it is quite possible that this observation arises because the experimentally defined average face is 'average enough', it is also possible that this observation reflects short-term adaptation to the statistics of the stimulus set. Moreover, one study reported norm-centred aftereffects for upright and inverted T-shapes despite the fact that there would be little reason to suspect that participants would have representations centred on such a seemingly arbitrary stimulus (e.g. Susilo et al., 2010a).

Experiment 8a provided a starting point for the exploration of adaptation aftereffects using complex visual stimuli, implementing a version of the task used by Leopold et al. (2001) using Greebles. While this task is not thought to reveal information about the direction of high-level aftereffects it has been used in previous studies to investigate face adaptation in adult prosopagnosic patients (Nishimura et al., 2010; Palermo et al., 2011), children (Nishimura et al., 2008; Nishimura, Robertson, & Maurer, 2011; Pimperton et al., 2009) and children with autism (Pellicano et al., 2007). Thus, this paradigm provides a starting point for investigating adaptation in novel objects.

The results of Experiment 8a were qualitatively consistent with the results of previous face adaptation studies that have used this paradigm. Specifically, adaptation to a computationally opposite Greeble facilitated recognition of the target Greebles more than adaptation to one of the non-opposite Greeble adaptors. Indeed, in line with the results reported by Leopold et al. (2001), there was a trend whereby non-opposite adaptors actually hindered recognition relative to baseline. Thus, it would seem that the results of studies using this paradigm to probe face recognition would need to be interpreted with caution. Specifically, it would be necessary to show more than a simple correspondence between the patterns of aftereffects in two populations of participants in order to conclude that they had similar face coding mechanisms.

As has been discussed, the paradigm implemented by Leopold et al. (used in Experiment 8a) is insufficient to detect whether aftereffects are centred on the average. With this in mind, Experiment 8b implemented an extended version of the paradigm used by Leopold et al. (2001) that has been used to rule out repulsive aftereffects in face adaptation, demonstrating that face aftereffects are centred on the central tendency of the population (though as demonstrated in Chapter 4, not necessarily discriminating norm and exemplar models). This task aims to investigate if aftereffects are centred on the central tendency of the population by increasing the distance of the adaptor from the average. If aftereffects are centred on the central tendency then adapting to the central tendency will cause no adaptation, with adaptation increasing as the adaptor is moved further away. On the contrary, if adaptation is centred on the adaptor then aftereffects will decrease as the adaptor moves further from the target. In the case of this study, adaptor centred effects would predict decreasing adaptation magnitude with distance from the average. Importantly, similar methods have been used to demonstrate the average-centred adaptation of identity (Leopold & Bondar, 2005; Susilo et al., 2010b), simple shapes (Susilo et al., 2010a) and emotion (Skinner & Benton, 2010) in adults, as well as typical norm-base coding in children and children (Jeffery et al., 2010). Thus, a similar finding for novel stimuli would be theoretically important.

In line with research into face adaptation, the results of Experiment 8b indicated that the magnitude of the aftereffects increased with distance from the target. Specifically, there was a significant main effect of adaptor distance from the average Greeble on threshold with paired t-tests revealing that both the -50% Greeble adaptor and the -100% Greeble adaptor resulted in significantly lower identification thresholds than the 0% (average Greeble) adaptor. These results would seem to suggest that, in line with studies into face adaptation, Greeble adaptation was centred on the central tendency of the population of Greebles used in this experiment. While in some sense this point (the average Greeble) is actually quite arbitrary, it would seem that the short term exposure to the population of Greebles was sufficient to tune the visual system to the population statistics of the Greeble stimuli.

Experiment 9 implemented a version of the paradigm used by Rhodes and Jeffery (2006) using Greebles. Following Rhodes and Jeffery, the magnitude of the aftereffect produced by ‘opposite’ and ‘non-opposite’ Greeble adaptors were compared, allowing the direction of aftereffects in Greeble-space to be inferred. That is to say, more adaptation for opposite than non-opposite trajectories (as has been found for face adaptation) would indicate that aftereffects biases perception towards the ‘opposite’ Greeble whereas equal adaptation for opposite and non-opposite trajectories would indicate that adaptation simply biased perception away from the adaptor Greeble. Contrary to the findings in the previous two studies, the results of Experiment 9 were not in line with the findings reported in the face adaptation literature. That is to say, there was no significant difference between the magnitude of the aftereffect measured following adaptation to an opposite Greeble adaptor and the magnitude of the aftereffect measured following adaptation to a non-opposite Greeble adaptor. Thus, while the studies reported in Chapter 5 do show some similarities between adaptation aftereffects in novel objects and the aftereffects reported in the face adaptation literature there might still be some important differences.

6.3 Future Directions and Unanswered Questions

6.3.1 Formalising the Norm-Based Model

Face-space models have proven to have substantial explanatory and predictive power, relying on the central assumption that, on each new viewing of a face, familiar or unfamiliar, the visual system makes use of the population statistics that have gathered through the observers experience with faces. A major theme underlying the research in this thesis, culminating in the attempt at formalising three versions of face-space in Chapter 4, has been the clarification of the many nuanced descriptions of possible face-spaces discussed in the literature, as well as their predictions on the various paradigms that have been put forward to differentiate them.

Perhaps most problematic in this regard is how to formalise descriptions of the norm-based model. For example, in its original conception, it was suggested that a suitable similarity metric for such a model might be the dot product between two exemplars (Valentine, 1991). However, the dot product would not seem to be suitable because, as Lewis (2004) has pointed out, it does not obey the rule that similarity is greatest when two vectors are identical. Specifically, given two vectors **a** and **b** the dot product is equal to the cosine of the angle between the two vectors, which is 1 when they are identical, multiplied by the magnitude of both vectors. Thus, one possibility is that similarity might be the normalised dot product, which is simply equivalent to the cosine of the angle. Indeed, this second suggestion is essentially the same as the implementation of the traditional norm-based model in Chapter 4 (e.g. Giese & Leopold, 2005). What is important to note here is that, unlike the norm-based model described by Valentine (1991), there is no role of exemplar density in this version of the norm-based model.

As was discussed in the previous section there does not seem to be a great deal of support for a traditional norm-based model. Indeed, research into lateral and oblique caricature (e.g. Lewis & Johnston, 1998; Lewis & Johnston, 1999; Rhodes et al., 1998, also see Chapter 1), as well as the research presented in Chapter 3 has provided evidence that is more in line with an exemplar-based account. However, one caveat is that, while a traditional norm-based model may not be in line with the data, there are an infinite number of possible similarity metrics that could be specified, making more or less use of the angle and Euclidean distance. Thus, while there may be little evidence that direction is the only relevant component of identity it would be difficult to rule out the possibility that it may play some role. In this sense it would be possible to imagine norm- and exemplar-based models as being separated by a continuum of possible models all making more or less use of direction information. Unfortunately at this point the debate becomes rather convoluted, making it difficult to see a clear way forward. Moreover, alternative versions of the norm-based model such as the two-pool model do not necessarily fit neatly into such a simple framework – though they do encode the location of the norm.

6.3.2 Adaptation in Different Categories of Faces

The findings reported in Chapter 4 open up the possibility that there could be an exemplar-based account of the effects seen in studies of face adaptation. Although Chapter 4 did provide evidence that an exemplar-based model can predict a wide range of face adaptation findings there are a number of paradigms that were not investigated. For example, a number of studies have suggested that there might be multiple 'norms' representing different race and gender faces (e.g. Jaquet, Rhodes, & Hayward, 2007; Jaquet et al., 2008; Rhodes et al., 2011). Indeed, it appears that it is actually possible to introduce dissociable aftereffects, biasing the perception of one category, say Black male faces, to appear to have expanded features (Convex appearance), while biasing the perception of a different category, say White male faces, to appear to have contracted features (Concave appearance).

It is possible that exemplar-based models might naturally explain such a finding (without having to appeal to the idea of multiple norms). As was discussed in Chapter 1, exemplar-based models directly reflect the overall statistics of a population of faces, with male and female faces, different race faces and other visually distinct groups represented by clusters in the face-space (Palmeri & Cottrell, 2010). In an exemplar-based model, adaptation to a face from one category, say a Black male face, would activate more Black male exemplars than adaptation to a different category, say White male. While it is difficult to say with any certainty without modelling such effects it seems likely that it would be possible to induce different directions of adaptation in two different categories simultaneously.

6.3.3 Adaptation in Familiar Faces

A second interesting avenue for future modelling work would be to look at the effect of familiarity on face adaptation. There have been several studies indicating that adaptation in familiar faces differs substantially from adaptation in unfamiliar faces (Carbon & Ditye, 2010; Carbon & Leder, 2005; Carbon et al., 2007;Laurence & Hole, 2011). Indeed, it has been noted that familiar face adaptation lasts much longer than unfamiliar face adaptation (maybe even longer than 24hrs). This finding is interesting with respect to the exemplar-based model presented in Chapter 4. Specifically, as was discussed it is generally the case that in the model the overall exemplar response of the space is relatively low when an unfamiliar face is presented. However, when a familiar face is presented the response of the corresponding exemplar is very high, possibly between 10 and 1000 times higher than the next most responsive exemplar (depending on the broadness of tuning). Thus, it is possible that the exemplar-based model would predict that adaptation to a familiar face would be longer lasting, taking more time to recover from than adaptation to an unfamiliar face. Indeed, using familiar faces as a probe might provide a way to differentiate norm- and exemplar-based models. This is because, while it is tempting to assume that the large adaptation associated with a familiar target exemplar would simply result in a strong adaptation aftereffect, some preliminary exploration of the model suggests that this is not the case. Rather, when the adapting face and test face are both the same familiar face the aftereffect appears to be centred away from the norm producing a local bias in perception away from the target location that is only slightly skewed towards an opposite face. However, as the dynamics of familiar face adaptation in the model appear to be very complex, and not easily intuited, it would require a full investigation of the model predictions in order to get an understanding of its behaviour. Nevertheless, research combining modelling work with behavioural data from studies of familiar and unfamiliar face adaptation would seem to be a very interesting topic for future research.

6.3.4 A Very Brief Comment on Neural Data

Although the focus of this thesis was very much on the behavioural data it is worth briefly considering how the findings fit in with the results from neurally-based studies (e.g. Davidenko, Remus, & Grill-Spector, In Press; Freiwald et al., 2009; Leopold et al., 2006; Loffler et al., 2005; Panis, Wagemans, & Op de beeck, 2011). Indeed, along with the face adaptation literature the results of neurally-based studies are frequently cited as providing strong support for a norm-based model of face recognition (e.g. Rhodes & Leopold, 2011; Susilo et al., 2010a; Susilo et al., 2010b). For example, using an fMRI adaptation paradigm, Loffler et al. demonstrated that average faces produced a lower fMRI BOLD signal change in the FFA than more distinctive faces. These results have been widely taken to support a norm-based model in which the neural population response is expected to be lower around the average face stimulus with the population response increasing for more distinctive faces (e.g. Anzures et al., 2009; Jeffery et al., 2010; Nishimura et al., 2008; Nishimura et al., 2011; Pimperton et al., 2009;

Rhodes & Jeffery, 2006; Tsao & Freiwald, 2006). In contrast, it is assumed that an exemplar-based model would predict that the neural population response would be highest for average face stimuli because this is where exemplars are most densely clustered and lowest for distinctive faces where exemplars are less densely clustered.

However, as Davidenko et al. (In Press) have pointed out there is a substantial flaw in the design of this study. Specifically, when measuring the response of a neural population to a particular stimulus set it is important to take account of the variability of the stimuli presented in each block. In Loffler et al. (2005) study typical faces were presented in one block followed by more distinctive faces in another block. Importantly as typical faces by their nature are all very similar (low variability) there will be a great deal of adaptation to the neural populations responding to these faces. However, as distinctive faces activate different neural populations (high variability) there will be substantially less adaptation carryover between trials in a given block. Indeed, repeating this experiment but controlling for the variability of stimuli within a block, Davidenko et al. actually demonstrated the opposite result. That is to say, there was a greater activation for typical than distinctive faces, in line with an exemplar-based model. Indeed, as was discussed in Chapter 5, it would seem that it is important to consider the short-term adaptive effects that can occur over the course of brief exposure to a particular stimulus set. Neural populations appear to adapt quickly to the statistics of a given set of stimuli, emphasising the response to the most distinctive stimuli and reducing the response to the least distinctive stimuli.

6.5 Final Conclusions

Differentiating norm- and exemplar-based models of face recognition has not proved to be straightforward, with both models making nearly identical predictions across a wide range of paradigms. However, with a slew of recent articles claiming to support a norm-based model (Anzures et al., 2009; Freiwald, Tsao, & Livingstone, 2009; Griffin, et al., 2011; Jeffery et al., 2010; Leopold & Bondar, 2005; Leopold et al., 2001; Nishimura et al., 2008; Nishimura et al., 2011; Pimperton et al., 2009; Rhodes, et al., 2011; Rhodes & Jeffery, 2006; Rhodes et al., 2010; Robbins, et al., 2007; Short et al., 2011; Susilo, et al., 2010a; Susilo, et al. 2010b; Tsao & Freiwald, 2006), as well as several recent reviews (e.g. Jeffery & Rhodes, 2011; Rhodes et al., 2005; Rhodes & Leopold, 2011; Rhodes & Jaquet, 2011) it is clear that a consensus has been reached. Importantly, the results reported in this thesis provide some challenge to this consensus, with the results presented in Chapter 4 suggesting that there may be alternative, exemplar-based accounts for the reported findings of studies into face adaptation. Furthermore, the results presented in Chapters 2 and 3, when taken together suggest that contrary to many claims (e.g. Blanz et al., 2000; Leopold et al., 2001; Rhodes et al., 2005; Tsao & Freiwald, 2006), direction may not be important in the representation of faces. Finally, Chapter 5 demonstrates that studies that use face adaptation paradigms, such as the ones reported by Leopold et al. and Leopold & Bondar (2005), to compare adaptive coding in children and adults may need to do more than simply demonstrate the qualitative similarity between the results. Thus, on balance the results in this thesis do not favour either an exemplar-based or a norm-based account; they do however suggest that the current consensus in favour of a norm-based model may be premature.

REFERENCES

Aguirre, G. K., Singh, R., & D'Esposito, M. (1999). Stimulus inversion and the responses of face and object-sensitive cortical areas. *NeuroReport*, 10, 189-194.

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

Anzures, G., Mondloch, C.J., & Lackner, C. (2009). Face adaptation and attractiveness aftereffects in 8-year olds and adults. *Child Development*, 80(1), 178-191.

Benson, P. J., & Perrett, D. I. (1991a). Synthesising continuous-tone caricatures. *Image & Vision Computing*, 9, 123-129.

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

Benson, P. J., & Perrett, D. I. (1994). Visual processing of facial distinctiveness. *Perception*, 23, 75-93.

Benton, A. L., Sivan, A. B., Hamsher, K. De S., Vamey, N. R., & Spreen, O. (1983). *Contribution to Neuropsychological Assessment*. New York: Oxford University Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.

Bilalic, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: Fusiform face area in chess experts and novices. *The Journal of Neuroscience*, 31, 10206-10214.

Blanz, V., O'Toole, A.J., Vetter, T., & Wild, H.A. (2000). On the other side of the mean: The perception of dissimilarity in human faces. *Perception*, 29, 885-891.

Brennan, S. E. (1985). The caricature generator. *Leonardo,* 18, 170-178.

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, 89, 60-64.

Burton, A. M., Bruce, V., & Dench, N. (1994). What's distinctive about a distinctive face? *The Quarterly Journal of Experimental Psychology*, 47A, 119-141.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, 23, 1-31.

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), *The oxford handbook of face perception* (pp. 287-306). New York: Oxford University Press.

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

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

Carbon, C. C., & Ditye, T. (2010). Sustained effects of adaptation on the perception of familiar faces. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 615-625.

Carbon, C. C., & Leder, H. (2005). Face adaptation: Changing stable representations of familiar faces within minutes? *Advances in Experimental Psychology*, 1, 1-7.

Carbon, C. C., Strobach, T., Langton, S. R. H., Harsanyi, G., Leder, H., & Kovacs, G. (2007). Adaptation effects of highly familiar faces: Immediate and long lasting. *Memory and Cognition*, 35, 1966-1976.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 335, 95-103.

Carey, S., Rhodes, G., Diamond, R., & Hamilton, J. (1992). Comparing the recognisability of caricatures, anticaricatures and lateral caricatures. Unpublished manuscript.

Craw, I. (1995). A manifold model of face and object recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (pp. 183-203). London: Routledge.

Dakin, S. C., & Omigie, D. (2009). Psychophysical evidence for a non-linear representation of facial identity. *Vision Research*, 49(18), 2285-2296.

Davidenko, N., Remus, D. A., & Grill-Spector, K. (In Press). Face-likeness and image variability drive response in human face-selective ventral regions. *Human Brain Mapping*.

Dennett, H., Edwards, M., & McKone, E. (2009). Are objects like faces? Norm-based versus exemplar-based coding as revealed by adaptation aftereffects. Abstract obtained from *Journal of Vision*, 9, 518a.

Duchaine, B., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, 41, 713-720.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge: MIT Press.

Fang, F., He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience*, 8, 1380-1385.

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

Frowd, C. D., Bruce, V., McIntyre, A., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P. J. B. (2006). Implementing holistic dimensions for a facial composite system. *Journal of Multimedia*, 1, 42-51.

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision Research*, 37, 1673-1682.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2, 568-573.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3, 191-197.

Gilaie-Dotan, S., & Malach, R. (2007). Sub-exemplar shape tuning in human face related areas. *Cerebral Cortex*, 17, 325-338.

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

Griffin, J. H., McOwan, W. P., Johnston, A. (2011). Relative faces: Encoding of Family resemblance relative to gender means in face space. Journal of Vision, 11(12), 1-11.

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

Hills, P. J., Holland, A. M., & Lewis, M. B. (2010). Aftereffects for face attributes with different natural variability: Children are more adaptable than adolescents. Cognitive Development, 25, 278-289.

Hintzman, D. L. (1990). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93, 411-428.

Hurlbert, A. (2001). Trading faces. *Nature Neuroscience*, 4, 3-5.

Jaquet, E., Rhodes, G., & Hayward, W. G. (2007). Opposite aftereffects for Chinese and Caucasian faces are selective for social category information and not just physical face differences. *The Quarterly Journal of Experimental Psychology*, 60, 1457-1467.

Jaquet, E., Rhodes, G., Hayward, W. G. (2008). Race-contingent aftereffects suggest distinct perceptual norms for different race faces. *Visual Cognition*, 16, 734-753.

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

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

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

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

Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition*, 4, 59-67.

Latinus, M., Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, 2, 1-12.

Laurence, S., & Hole, G. (2011). The effect of familiarity on face adaptation. *Perception*, 40, 450-463.

Leopold, D. A., & Bondar, I. (2005). Adaptation to complex visual patterns in humans and monkeys. In C. W. G. Clifford, & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and after-effects in high-level vision* (pp. 213-240). Oxford: Oxford University Press.

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

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4(1), 9-94.

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

Lewis, M. B., & Johnston, R. A. (1998). Understanding caricatures of faces. *Quarterly Journal of Experimental Psychology*, 51A, 321-346.

Lewis, M. B. & Johnston, R. A. (1999a). A unified account of the effects of caricaturing faces. *Visual Cognition*, 6, 1-41.

Lewis, M. B. & Johnston, R. A. (1999b). Are caricatures special? Evidence of peak shift in face recognition. *European Journal of Cognitive Psychology*, 11, 105-117.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 212-228.

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

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

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Megreya, M., & Burton, A. M. (2006) Unfamiliar faces are not faces: Evidence from a matching task. *Memory and Cognition*, 34, 865-876.

Meytlis, M. (2011). A model of face space. *Visual Cognition*, 19, 13-26.

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

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

Nishimura, M., Robertson, C., & Maurer, D. (2011). Effect of adaptor duration on 8 year olds' facial identity aftereffects suggests adult-like plasticity of the face norm. *Vision Research*, 51, 1216-1222. Forthcoming.

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

Palermo, R., Rivolta, D., Wilson, C.E., & Jeffery, L. (2011). Adaptive face space coding in congenital prosopagnosia: Typical figural aftereffects but abnormal identity aftereffects. Neuropsychologia.

Palmeri, T. J., & Cottrell, G. W. (2010). Modeling perceptual expertise. In I. Gauthier, M. J. Tarr, & D. Bub (Eds.), *Perceptual expertise: Bridging brain and behavior* (pp. 197-244). New York: Oxford University Press.

Panis, S., Wagemans, J., & Op de Beeck, H. P. (2011). Dynamic norm-based encoding for unfamiliar shapes in the human visual cortex. *Journal of Cognitive Neuroscience,* 23, 1829-1843.

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

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

Pothos, E. M., & Wills, A. J. (2011). Introduction. In E. M. Pothos, & A. J., Wills (Eds.), *Formal approaches in categorization* (pp. 1-17). Cambridge: Cambridge University Press.

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

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

Rhodes, G. (1996). *Superportraits: Caricatures and recognition*. Hove, UK: Psychology Press.

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

Rhodes, G., Carey, S., Byatt, G., & Proffitt, F. (1998). Coding spatial variations in faces

and simple shapes: a test of two models. *Vision Research*, 38, 2307-2321.

Rhodes, G., & Jaquet, E. (2011). Aftereffects reveal that adaptive face-coding mechanisms are selective for race and sex. In R. A. Adams Jr., N. Amabady, K. Nakayama & S. Shimojo (Eds.), *The science of social vision*. New York: Oxford University Press.

Rhodes, G., Jaquet, E., Jeffery, L., Evangelista, E., Kean, J., & Calder, A.J. (2011). Sex specific norms code face identity. *Journal of Vision*, 11(1), 1-11

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

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

Rhodes, G., & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. W. Calder, G. Rhodes, M. H. Johnston, & J. V. Haxby (Eds.), *Oxford handbook of face perception*. Oxford: Oxford University Press.

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

Rhodes, G., & McLean, I. G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: towards a model of configural coding. *Perception*, 19, 773-794.

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

Rhodes, G., & Tremewan, T. (1994). Understanding face recognition: Caricature effects, inversion and the homogeneity problem. *Visual Cognition*, 1, 275-311.

Rhodes, G., Watson, T. L., Jeffery, L., & Clifford, C. W. G. (2010). Perceptual adaptation helps us identify faces. *Vision Research*, 50, 963-968.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2010). Inverted faces are (eventually) processed holistically. *Vision Research*, 51, 333-342.

Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? Cognition, 88, 79-107.

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

Short, L. A., Hatry, A. J., & Mondloch, C. J. (2011). The development of norm-based coding and race-specific face prototypes: An examination of 5- and 8-year olds' face space*. Journal of Experimental Child Psychology*, 108,338-357.

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

Susilo, T., McKone, E., & Edwards, M. (2010a). Solving the upside-down puzzle: Why do upright and inverted face aftereffects look alike? *Journal of Vision*, 10(13), 1-16.

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

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

Stevenage, S. V. (1997). Face facts: Theories and findings. *The Psychologist*, 10, 163-168.

Tanaka, J. W., & Corneille, O. (2007). Typicality effects in face and object perception: Further evidence for the attractor field model. Attention, Perception & Psychophysics, 69, 619-627.

Tanner, T. G., Hill, N. J., Rasmussen, C. E., Wichmann, F. A. (2005). Efficient adaptive sampling of the psychometric function by maximizing information gain. In H. H. Bulthoff, A. Mallot, R. Ulrich, & F. A. Wichmann (Eds.). *Proceedings of the 8th Tubinger Perception Conference* (Vol. 106). Kirchentellisfurt: Knirsh Verlag.

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

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

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *Computer Graphics and Applications IEEE*, 21, 42-50.

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

Valentine, T. (2001). Face-space models of face recognition. In M. J. Wenger, & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 83-114). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognizing and classifying faces. *Perception*, 15, 525-535.

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

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

Wickham, L. H. V., Morris, P. E., & Fritz, C. O. (2000). Facial distinctiveness: Its measurement, distribution and influence on immediate and delayed recognition. *British Journal of Psychology*, 91, 99-123.

Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research*, 42, 2909-2923.

Wong, A. C. N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert – but which type? *Psychological Science*, 20, 1108-1117.

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

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

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