THE ROLE OF SYNCHRONY IN PERCEPTUAL BINDING OF FACIAL MOVEMENT

Ben Brown

Submitted to the School of Psychology
of the University of Nottingham
for the degree of Doctor of Philosophy

September 2017

ABSTRACT

Faces are some of the most complex and important sources of information in the visual world. Their motion transmits a rich stream of social signals that human observers nonetheless seem able to interpret effortlessly. The aim of this thesis is to investigate how the timing of coordinated facial movement is represented by the visual system.

Chapter 1 reviews the current knowledge of several major themes in the face perception literature. It first reviews three classic paradigms that each have a bearing on the present work, before introducing the theoretical issues to which they have most often been applied: holistic face processing and expertise. This chapter then reviews the major models of face processing and its neural implementation, before turning to expressive or dynamic faces, and finally recent reports of timing-mediated cross-feature interactions.

Chapter 2 begins by describing the overall stimulus and paradigm (many details of which apply across behavioural experiments), before presenting the first three experiments. Together, these confirm that perception of one feature's motion is not independent of concurrent motion in another feature, despite the latter being spatially separate and task-irrelevant.

Chapter 3 presents Experiments 4 and 5, which reveal the specificity of these interdependencies to certain dynamic feature relationships (particularly synchrony) and the right visual hemifield, respectively.

Chapter 4 presents Experiments 6 and 7, which employ visual adaptation to probe the neural encoding of relative feature timing. Experiment 6 discounts an orientation-like channel-based account, instead finding a general reduction in performance following adaptation to asynchronous feature movement. Experiment 7 discounts recalibration of perceptual models centred on synchrony as an explanation.

Chapter 5 presents a combined dynamic localiser and phase-encoded fMRI experiment, which found greater modulation by relative feature timing of facial motion sensitive brain areas. Using a simple contrastive analysis, this experiment provides the first evidence that regions involved in facial motion processing prefer certain feature relative timings.

Chapter 6 concludes this thesis by reviewing, contextualising, and critiquing the present work. Included is a summary of the original contribution to knowledge, and potential future directions.

ACKNOWLEDGEMENTS

I would like to thank my primary supervisor, Alan Johnston, for his patience and for all the meetings that I entered dispirited but left inspired. His enthusiasm for my work has been genuine and unwavering. I would also like to thank my secondary supervisors, Marty Sereno and Denis Schluppeck, for their invaluable support and willingness to induct me into the practicalities of brain imaging.

I am grateful to my my parents, my sister, and my brothers. I have enjoyed many heated but good humoured debates with them, and hope to enjoy many more.

I would also like to thank my closest friends Elliot, Luc, Matt, and Ollie, with whom I have had some of my deepest conversations and messiest evenings.

Finally, I would like to thank my partner Leanne. We are a team.

CONTENTS

ABSTRACT **1**

ACKNOWLEDGEMENTS **2**

CONTENTS **3**

CHAPTER 1. Introduction **5**
1.1. Overview *5*
1.2. Classic experimental signatures of face perception *5*
1.3. Holistic face processing and the expertise hypothesis *17*
1.4. Models and neuroanatomy of face processing *22*
1.5. Representing faces: spaces, norms and exemplars *26*
1.6. Expressions and dynamic face perception *32*
1.7. Temporal relationships and cross-feature interactions *39*

CHAPTER 2. Dynamic features are not processed independently **42**
2.1. Introduction *42*
2.2. General Methods *42*
2.3. Experiment 1: Adaptation to fast facial motion causes illusory slowing *51*
2.3.1. Introduction 51
2.3.2. Methods 52
2.3.3. Results 55
2.3.4. Discussion 56
2.4. Experiment 2: Mouth movement impairs eyebrow discriminability *57*
2.4.1. Introduction 57
2.4.2. Methods 58
2.4.3. Results 62
2.4.4. Discussion 63
2.5. Experiment 3: De-correlating eyebrow and mouth movement abolishes mouth movement interference on eyebrow discriminability *63*
2.5.1. Introduction 63
2.5.2. Methods 64
2.5.3. Results 66
2.5.4. Discussion 67

CHAPTER 3. Dynamic feature binding is timing-dependent, and this dependency is lateralised **69**
3.1. Introduction *69*
3.2. Experiment 4: Eyebrow-mouth perceptual interactions are dependent on relative timing *70*
3.2.1. Introduction 70
3.2.2. Methods 72
3.2.3. Results 75
3.2.4. Discussion 81
3.3. Experiment 5: The timing dependency of eyebrow-mouth perceptual interactions is specific to the right visual hemifield *83*
3.3.1. Introduction 83
3.3.2. Methods 86
3.3.3. Results 89
3.3.4. Discussion 91

CHAPTER 4. Adaptation to facial motion reveals the importance of synchrony **92**
4.1. Introduction *92*

4.2. Experiment 6: Adaptation to asynchronous facial motion impairs eyebrow discriminability irrespective of relative timing at test ... *93*
4.2.1. Introduction ... 93
4.2.2. Methods ... 95
4.2.3. Results ... 98
4.2.4. Discussion ... 100
4.3. Experiment 7: Little evidence for adaptive recalibration of internal models of facial timing centred on synchrony ... *103*
4.3.1. Introduction ... 103
4.3.2. Methods ... 106
4.3.3. Results ... 108
4.3.4. Discussion ... 112

CHAPTER 5. Facial motion sensitive brain areas are modulated by relative feature timing ... **114**
5.1. Introduction ... *114*
5.2. Methods ... *120*
5.2.1. Participants ... 120
5.2.2. Scanning apparatus and parameters ... 121
5.2.3. Dynamic localiser ... 121
5.2.4. Phase-encoded design ... 123
5.3. Results ... *124*
5.4. Discussion ... *131*

CHAPTER 6. Conclusion ... **133**
6.1. Overview ... *133*
6.2. Summary of results ... *133*
6.3. Contribution to knowledge and theoretical implications ... *135*
6.4. Limitations of present work and future directions ... *137*
6.5. Conclusion ... *144*

REFERENCES ... **145**

CHAPTER 1. Introduction

1.1. Overview

This general introduction overviews the key concepts of the face perception literature. Additional studies of specific relevance to our experiments are introduced in those experiments' sections below; in particular the neuroscience of facial motion processing is considered in greater depth in Chapter 5.

This general introduction presents three classic paradigms of face perception research, the results and methodological issues of each being of relevance to the present work. Holistic processing is introduced and its potential specificity for faces discussed. We then describe the major dual-stream models of face processing, what these processes may be doing (contextualised in terms of broader object recognition), and how they may be implemented as neural spaces by norm-based coding. The additional challenges of facial motion research are then discussed, distinctions between basic and more nuanced conversational expressions made, and current understanding of information transmitted in the dynamics of the facial motion stream described. Finally, we review a recent investigation into the role of timing in dynamic feature binding, from which the present work takes its starting point.

1.2. Classic experimental signatures of face perception

A number of behavioural results suggest that face perception is unlike other forms of visual recognition. Here we introduce three classic perceptual effects, upon which the claim of many experiments (including our own) to be tapping face-specific visual mechanisms is based.

An inversion effect is a reduction in performance (recognition memory or perceptual matching) when stimuli are presented upside-down (McKone & Robbins, 2011). Both recognition and matching typically entail within-class discrimination: ascertaining whether two images depict the same object as opposed to merely the same *kind* of object. Clearly an inversion effect can only manifest for objects possessing a canonical upright, but amongst these

reductions in performance following inversion are not uncommon (Robbins & McKone, 2007).

Yin (1969) was first to report a *disproportionate* inversion effect for faces. During a learning phase, participants serially viewed 40 images depicting faces, houses, airplanes or men in motion (two such classes per block). During a subsequent test phase, they then had to identify which of several pairs of test images had previously been seen. Four such blocks were completed per participant: two presenting upright images and two inverted. Thus, a greater number of images erroneously identified as belonging to the learning set in the inverted condition (relative to upright) constituted an inversion effect.

Yin found that the error rate was not only lowest for faces in the upright condition, but also highest in the inverted condition. He therefore concluded that difficulty with faces following inversion reflected disruption of face-specific mechanisms, over and above more general impairment of object recognition. Like Galton (1883, as cited in Tanaka & Simonyi, 2016) before him, Yin speculated that recognition entailed "a general impression of the whole picture", and inversion subsequently became widely used as a "tried and true" measure of holistic face processing (Tanaka & Simonyi, 2016).

Initial results suggested that the inversion effect was not present in children, suggesting that the specialisation for face processing, as indexed by the inversion effect, takes time to develop. Carey & Diamond (1977) presented children aged 6, 8 and 10 years with training sets of upright and inverted faces and houses. Test pairs containing a previously shown and new item were then presented (with the same orientation and stimulus type), to assess recognition memory. They found no disproportionate inversion effect for faces over houses in participants younger than 10 years. A second experiment assessed the possibility that differing reliance on feature-wise versus holistic processing was the cause. They presented children again aged 6, 8, and 10 years with faces whose expressions and 'paraphernalia' (e.g. clothing, hairstyle, glasses) differed, the task being immediate recognition of an individual from a test pair. Performance improved with age as predicted, consistent with decreased dependence on the 'piecemeal' cues for identification by older children.

This early evidence was rapidly refuted, however (McKone & Robbins, 2011). In a review, Crookes & McKone (2009) argued that the literature favoured general cognitive development as the source of performance improvement through childhood, with the qualitative aspects of performance (and therefore face-specific perceptual mechanisms) present from at least 5 – 7 years. They argue that numerical differences in performance over age are not easily comparable due to floor effects. Notably, inversion effects present in older children but not younger were shown to be a product of older children's higher performance in the upright conditions, versus the restricted range of responses (at floor) of younger children across conditions. Studies without such range restriction reported a stable inversion effect with age.

The Thatcher effect (P. Thompson, 1980) is a striking example of differential perceptual ability for upright and inverted faces. Using an image of Margaret Thatcher, Thompson inverted the eyes and mouth, noting that whereas this manipulation was immediately obvious in upright faces (making them appear grotesque), it was less detectable within inverted faces (Figure 1.1). Stürzel & Spillmann (2000) asked participants to judge when Thatcherised faces switched between appearing 'pleasant' and 'grotesque' while rotating them between upright and inverted orientations, the logic being that a Thatcherised face starts to appear grotesque when the stimuli comes close enough to upright to render holistic processing possible. They found the changeover to occur within a fairly narrow range of orientations 94 – 100° from the vertical (in either direction), suggesting a step-change from componential to holistic processing at this point.

Thompson, P. (1980). Margaret Thatcher: A New Illusion. Perception, 9(4), 483–484.

Figure 1

Figure 1.1. The Thatcher illusion. Manipulation of the image on the right is much more obvious when the faces are presented upright. From Thompson (1980).

The argument that inverted faces are processed in a qualitatively different manner is not without scrutiny. Contrasting with Stürzel & Spillmann (2000), Lewis (2001) reported a gradual loss of configural information with increased rotation from upright when testing reaction time for detection of Thatcherisation. Sekuler, Gaspar, Gold, & Bennett (2004) presented participants with single target faces obscured by additive Gaussian noise, which participants had to match from subsequent pairs of unobscured test faces. By considering trial-by-trial performance while taking account of which parts of the target face the random Gaussian noise had obscured, they were able to estimate which parts of the face were informative for identification. This was repeated with target and

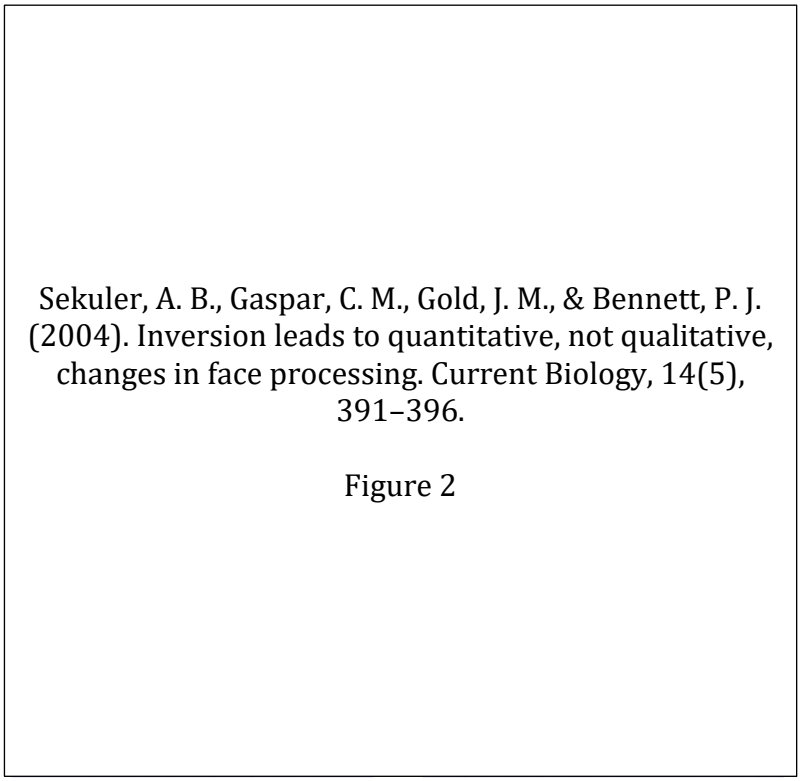


Figure 1.2. Classification images for male and female (top, bottom) upright and inverted (left, right; latter after rotation) faces. There is little difference between orientations. From Sekuler et al. (2004)

test faces upright or inverted. Face contrast (which renders faces more visible against the noise) had to be substantially higher for inverted faces to attain performance parity with upright faces, as expected. However, the regions of the face found to be informative for the matching decision did not differ across upright and inverted faces; rather, comparison against a simple ideal observer suggested information from those regions was used more efficiently in the upright case (Figure 1.2).

Rossion (2008) did not find this quantitative (as opposed to the prevailing qualitative) conception of the inversion effect convincing. With reference to Sekuler et al. (2004) specifically, Rossion argues that finding the same areas of the face to be informative for upright faces as for inverted faces does not speak to whether those areas are analysed in a piecemeal or configural fashion. He also suggests that Sekuler et al.'s data does in fact show signs of configural processing that is greater for upright faces, in terms of the positioning and precision of the pixels found to be informative. More generally, Rossion argues that studies claiming to provide evidence against differential configural processing for upright versus inverted faces do not replicate the studies that they contradict, and furthermore proposes that loss of holistic representation and disruption to configural processing are conceptually, but not experimentally, distinguishable.

Despite the ubiquity of inversion tasks (including within our first five experiments), reduced performance with inversion now tends to be regarded as an indirect measure (McKone, 2010). This is because inversion does not directly manipulate most aspects of faces believed to be important for holistic processing such as distances between features, and because the existence of (smaller) inversion effects for non-face objects sustains the possibility that both stimuli are processed by the same mechanisms (albeit with differing efficiency). Which aspects of the stimulus are in fact relevant to holistic processing forms a related debate that we consider below. For now it suffices to say that Yin's original finding of the exceptional nature of face inversion has survived considerable argument to the contrary, and is substantiated by a large body of research (McKone & Robbins, 2011; Robbins & McKone, 2007). Furthermore, that inversion preserves more aspects of the original stimuli than alternative manipulations (e.g. contrast polarity reversal or horizontal splitting) is advantageous, because it minimises confounds when one's goal is to dissociate face-specific from generic- or lower-level visual mechanisms.

Because inversion (and the composite and part-whole paradigms) target face processing, we have necessarily invoked concepts such as 'holistic', 'configural', and 'feature-wise' prior to their explication in the next section. For the purposes of this thesis we operationalise 'holistic processing' as an inability to consider

single elements independently, such that one obligatorily impinges on perception of another. Such perceptual interdependencies imply joint-coding of features, which may occur via 'configural processing': the encoding of a face by the inter-feature distances that determine its configuration. 'Feature-wise' or 'piecemeal' processing refers conversely to consideration of each feature separately, for example serially. Inversion is widely used to dissociate face-specific and holistic from generic and feature-wise processing despite its indirect relationship to these because of the much greater magnitude of the inversion effect for faces over other objects (see above), and because of convergent evidence from more direct manipulations tested in combination with inversion (see below). These and the desirable conservation of lower-level visual properties under inversion justify our use of this manipulation in our first five experiments.

Although the apparent face-specificity of both the composite and part-whole effects is tied to their modulation by inversion, they are regarded as more direct measures of configural processing because they manipulate the interrelation of features and whole face context. The composite effect is a reduction in performance (recognition memory or matching) when judging the top half of stimuli whose bottom halves have been replaced by that of another stimulus from the same class. For faces, this amounts to greater difficulty in identifying the top half when combined with another face's bottom half (recognition), or correctly determining that two such composite faces share the same top halves when their bottom halves differ (matching). Thus, although only the top halves (and features therein) are relevant to the task, the conjoined bottom halves nonetheless impinge on performance (Rossion, 2013).

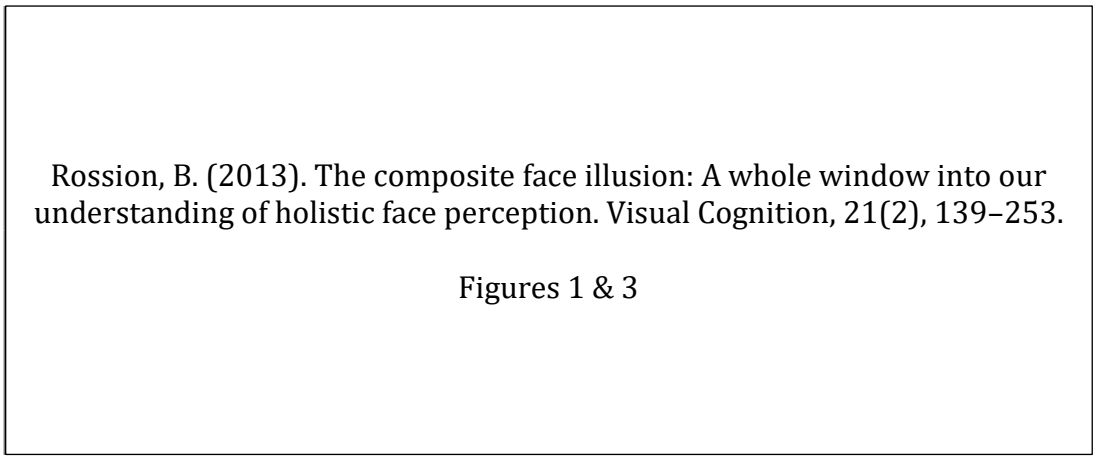


Figure 1.3. The composite face effect. All face top halves are identical, but they appear different when paired with differing bottom halves (top row). The effect is abolished with horizontal misalignment of the halves (bottom row). From Rossion (2013).

In first reporting the effect, Young, Hellawell & Hay (1987) demonstrated two manipulations which release this interference: horizontal misalignment of the incongruous halves, and inversion. The former entailed recognition of famous face halves and non-famous top face halves that had been pre-learned (Experiments 1 and 3 respectively). While error rates were typically under 5%, Young et al. found reaction times for correct responses were slower when the two halves were offset horizontally so as to break up the 'face shape'. This effect persisted even when participants had previously learned the half to be identified in isolation (Experiment 3). Importantly, the difference in reaction times between aligned versus misaligned face half conditions disappeared when the faces were inverted (Experiment 2), representing a performance *improvement* following inversion for this explicitly non-holistic judgement. The authors concluded that "perception of a novel configuration in facial composites interferes with recognition of their constituent parts", and that inversion worked to reduce the (here detrimental) perception of facial configuration.

Hole (1994) introduced the matching-based composite face task, which unlike recognition does not require prior familiarity and is therefore now the significantly more popular design (Rossion, 2013). Participants view two composite test faces (either simultaneously as per Hole, 1994, or more often sequentially, Rossion, 2013) and have to judge whether the top halves belong to the same or different individuals. Reaction times are longer when the top halves are aligned with different bottom halves, again suggesting this makes identical

top halves appear different. Hole compared performance with upright faces to inverted, and similarly to Young, Hellawell, & Hay (1987) found better performance (as measured by reaction time) for the latter.

One contention of this design surrounds which combinations of same and different top and bottom halves should be included. Typically, trials present a mix of top-half-same and top-half-different trials, but only the former are included in analysis. For example, the simplest measure of the effect is how many top-half-same composite face pairs are erroneously judged to be different, when their (always differing) bottom halves are aligned versus when they are misaligned. Gauthier and colleagues (Cheung, Richler, Palmeri, & Gauthier, 2008; Gauthier & Bukach, 2007; Richler, Cheung, & Gauthier, 2011) refer to this as a 'partial' design in contrast to their favoured 'complete' design, which incorporates top-half-different and bottom-half-same trials (Figure 1.4).

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? Cognition, 103(2), 322–330.

Figure 1

Figure 1.4. All possible combinations of same/different top and bottom halves comprising the 'complete' design. Only the incongruent same and congruent different composites are analysed in the 'partial' design. Letters denote face identities, +/- denote whether face halves are the same or different. From Gauthier & Bukach (2007).

Using this design, Gauthier and colleagues challenged several commonly made arguments regarding the nature and specificity of holistic face processing (considered below). Their position was that in order to assess the impact of the irrelevant bottom face halves on matching of the top halves, comparison of 'congruent' versus 'incongruent' trials must be made – that is, trials where the top and bottom face halves would indicate the same response versus conflicting responses. A lack of difference in response pattern between these two types of trial would argue against the response indicated by the irrelevant face half having any bearing on participants' decision. There are two ways in which a trial could be congruous (with both halves of the faces either the same or different),

and similarly two forms of incongruous trial (where one half of the faces indicates a 'same' response while the other indicates 'different'). Gauthier and colleagues pointed out that the common 'partial' design entails presentation of only incongruous 'same' trials and congruous 'different' trials, with analysis of just the former constituting consideration only of 'hit rates' in signal detection theory parlance (Cheung et al., 2008). They therefore considered the partial design susceptible to confound, unable to distinguish discriminability from response bias, and liable to mislead.

In a lengthy review, Rossion (2013) defended the classic composite face design and criticised the 'complete' variant. He argued that the casting of holistic processing in attentional rather than Gestalt terms was inappropriate, with only the former justifying the congruent-incongruent dichotomy. Specifically, he argued that differential performance based on similarity of the irrelevant half cannot be taken as evidence for its integration with the half to be judged (holistic binding), but rather of potentially generic interference only. The deployment of an attentional distraction-based rationale for measurement of perceptual integration therefore lacks construct validity. He also defended the selective analysis of top-half-same trials by appealing to the fact that it is easier to make two identical stimuli look different than vice-versa: there are innumerable ways by which stimuli may appear different yet only one need be detected to correctly make that judgement; by contrast an incorrect 'same' judgement requires perceptual nulling of all physical stimulus differences exactly. Therefore the lack of effect found when composing same bottom halves with different top halves (i.e. that this does not make different top halves illusorily appear more similar) need not pertain to measurement of the overall composite face effect and may be legitimately excluded from analysis.

Overall, like the inversion effect the relation of the composite face effect to perceptual processing is not without controversy, but it remains a widely deployed paradigm. Whereas the disproportionate inversion effect suggests a dependence on higher-order spatial properties (orientation) that is specific to faces over other objects, the composite face effect implicates obligatory

perceptual binding across facial features that persists even when piecewise analysis would improve performance.

Our final illustrative paradigm is the part-whole task. Like the original composite face task, it entails recognition from memory but this time of individual features in differing whole-face face contexts. Tanaka & Farah (1993; Experiment 1) presented participants with whole faces in a learning phase, before testing them on recognition of single features in a subsequent test phase. The target and distractor features were presented in isolation (e.g. just noses with the question "Which is Larry's nose?") or in identical whole-face contexts. That is, even in the whole-face test condition the target and distractor did not differ except by a single feature. This procedure was repeated with feature-position-scrambled learning and test whole faces. An improvement in performance for recognising features in whole face contexts over those presented in isolation would be noteworthy given said context did not add any discriminative information.

Tanaka & Farah found such a facilitative effect of whole-face context only for non-scrambled faces. The restriction of the effect to intact faces argues against a general 'encoding specificity' advantage of recall in the same context as learning, in favour of a specialised representation of facial features which depends on their typical positioning within a face (see also Experiment 3's discussion below).

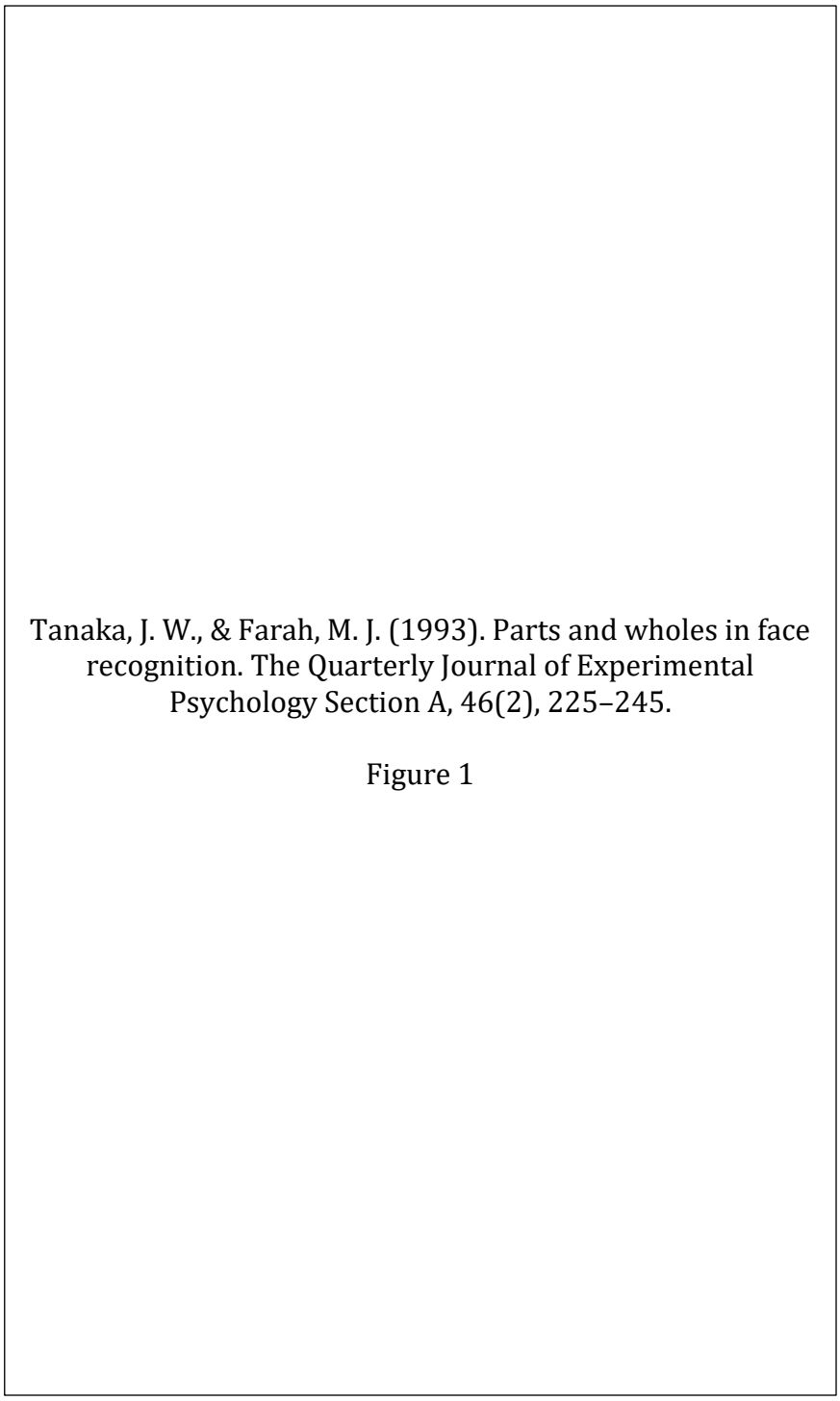


Figure 1.5. Tanaka & Farah's (1993) three test stimulus types from their experiment 1. Participants had to discriminate a feature of a previously studied face in isolation, within identical whole face contexts, or identical scrambled face contexts.

Tanaka & Sengco (1997) repeated the experiment but this time trained participants with faces in which the inter-eye distance was manipulated. They tested feature recognition in whole face contexts where the inter-eye distance either matched or differed from the learning phase. Isolated feature recognition (no face context) was again worst, replicating their earlier result. Furthermore, feature recognition in whole face contexts that matched the learning phase was better than for those whose inter-eye distance differed. Notably, this was the case for recognition of the nose and mouth as well as the manipulated eyes, suggesting that the latter was incorporated in the encoding of the former despite not being relevant to the task. In other words, representation of individual features is dependent on configuration elsewhere in the face. No such configurational interdependency was found for inverted faces or houses, however.

It is natural to consider what the result of training on isolated features would be. Leder & Carbon (2005) tested many combinations of isolated parts and wholes, presented at training and test, as well as testing at both orientations. Following training on features in isolation, testing their recognition within a whole face context actually worsened performance, and this remained the case even when participants were informed in advance as to the feature that would be judged. It would appear that the effect of face context on feature recognition helps or hinders dependent on training regime, with the novel whole-face context at test necessarily interfering in the above case.

Overall then, like the composite face effect the part-whole effect seems to be tapping obligatory binding mechanisms. Like the inversion effect, it may improve or degrade performance dependent on the role of whole faces in the design. Unlike the two previous paradigms however, it does require a greater degree of stimulus manipulation – removing or swapping elements as opposed to merely splitting or inverting. This aside, the part-whole task has not elicited the same debate over procedural or theoretical ambiguity.

The underlying mechanisms of face perception are discussed below, but the three classic tasks above introduce key themes that are of relevance across our experiments. Firstly, these mechanisms appear highly contingent on orientation.

Thus, although the visual system is capable of high performance through specialisation, this is readily 'disabled' by a comparatively appearance-preserving stimulus manipulation. This inflexibility is not only suggestive of processing strategy (e.g. initial 'template matching') but provides an invaluable experimental control in inverted stimuli.

Secondly, these mechanisms appear obligatory and are engaged even when detrimental to task performance. Again automaticity may be a cost of specialisation, but the ability of irrelevant face information to interfere with judgements is another useful property central to our experiments.

Thirdly, these mechanisms appear to process features together, so that the appearance of one region of the face is dependent on the others. This joint processing of features and/or indivisibility of the overall face representation is core to most descriptions of face perception, extension of which into the dynamic domain was the primary goal of our experiments.

1.3. Holistic face processing and the expertise hypothesis

The issues associated with the inversion and composite face effects relate to what assumptions are made regarding the underlying phenomena of interest. Various definitions of 'holistic' have been applied (Maurer, Le Grand, & Mondloch, 2002; Piepers & Robbins, 2012; Richler & Gauthier, 2014), broadly segregable into those emphasising 'first-order' or 'second-order' relations between features. Note then that our earlier operationalisation of holistic processing (obligatory perceptual interdependencies between features) accommodates both definitions.

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. Frontiers in Psychology, 3(DEC), 1–11.

Figure 4

Figure 1.6. Different meanings of configural representation behave differently under transformation, dependent on which distances are measured. Notice the definition based on fiducial points positioned at feature centres (right) is invariant to mouth shape. From Piepers & Robbins (2012).

Penev, P. S., & Sirovich, L. (2000). The global dimensionality of face space. In Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, March, 264–270.

Figure 1

Figure 1.7. The first 16 principal components plus 8 higher ones produced by PCA on faces (eigenfaces) from the FERET dataset. Because PCA produces dimensions ordered by the amount of image variance explained, the first few components represent broad qualities such as lighting. Lower components represent facial appearance at increasingly fine spatial scale. From Penev & Sirovich (2000).

The first-order relational properties of a face are its overall structure. Goffaux & Rossion (2006) found a dependence of holistic processing on low spatial frequencies present in facial stimuli, wherein finer featural information is absent. It is notable that principal components analysis (hereafter PCA) on faces as described below delivers components representing overall appearance at increasing levels of fine detail, rather than segregated features (Figure 1.7). In other words, decomposing faces according to their natural variability yields a space whose dimensions represent global, not local structure (Draper, Baek, Bartlett, & Beveridge, 2003). This emphasis on overall structure also concords with arguments for an initial gating detection stage in the visual system, whereby subsequent face-specialised cortical machinery is only engaged when a coarse face-shape template is matched to visual input (Tsao & Livingstone, 2008). Face inversion is then believed to defeat this template matching, forcing regression to generic mechanisms. By this account first-order properties therefore transcend specific individuals, instead being an aspect common to all faces.

Second-order relational properties are the inter-feature distances subject to configural processing (defined earlier), and 'configural' and 'holistic' processing are often considered synonymous. Even while accepting the configural meaning however, differences remain in measuring these relationships. Distances measured from fiducial points in the centre of features will differ from those positioned at features' outlines and this has consequences when faces vary, such as mouth movement during speech (Piepers & Robbins, 2012; Figure 1.6). Nonetheless, second-order properties offer a means by which individuation of one face from another may be achieved.

A larger controversy concerns the 'special' nature of face perception – whether the apparent cortical specialisations and exceptional behavioural results reflect expertise rather than innate ability. The expertise hypothesis argues that any class of similar objects requiring individuation will lead to development of similar cognitive strategies and deployment of equivalent cortical processing, and originates in Diamond & Carey's (1986) report of an inversion effect for dogs in dog experts. Subsequent attempts to replicate this have been mixed, however

(Kanwisher & Yovel, 2006). Gauthier and colleagues' experiments with *Greebles* form a substantial part of the argument (Gauthier, Behrmann, & Tarr, 1999; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Greebles are computer-generated 3D shapes intended to contain similar cues for individuation as faces, but otherwise be artificial and non-suggestive of real objects (Figure 1.8). Similar figural effects, fusiform face area (FFA, see below) activity, electrophysiological response and impairment in prosopagnosia have been reported for Greebles as for faces (Tarr & Cheng, 2003), leading the researchers to posit generalised rather than face-specific mechanisms.

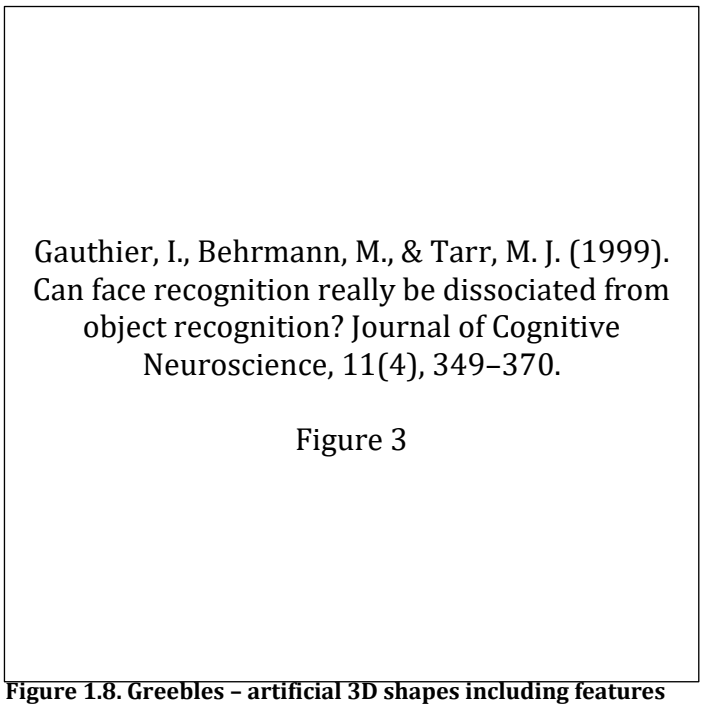


Figure 1.8. Greebles – artificial 3D shapes including features with which they can be individuated. From Gauthier et al. (1999).

Again however, these findings are contested (Gauthier & Bukach, 2007; McKone & Robbins, 2007; Robbins & McKone, 2007). Gauthier and colleagues reported face-like effects (such as the composite effect) for Greebles following only hours of training, so the expertise that they argue underlies performance for faces and Greebles alike must accumulate rapidly as opposed to over a lifetime (Duchaine, Dingle, Butterworth, & Nakayama, 2004). Duchaine et al. tested a prosopagnosic patient "Edward" with a standard Greebles task, entailing progressive learning of Greeble 'families' and names over several sessions (ability to correctly identify

the families versus names of Greebles is intended to test differing levels of individuation). If Greeble processing is equivalent to face processing in that both depend on rapidly attainable expertise, then a patient lacking such an ability for faces should also fail to learn Greebles. Edward's performance over training sessions did not differ from controls on any measure however – neither percent correct nor reaction time. The ability of a severe prosopagnosic to nonetheless learn Greebles as well as non-prosopagnosics argues against their processing by common mechanisms, and more generally that short-term Greeble training is not an effective test of expertise.

Central to the expertise argument is the attentional and congruence-based conception of the composite face effect described above, which can show differences between novices and experts for non-face object categories such as cars (Gauthier, Curran, Curby, & Collins, 2003). McKone & Robbins (2011) argue that equating holistic binding to failure of selective attention in this way is akin to arguing that ink colour and word identity are perceptually integrated in the Stroop effect, though this is known not to be the case. To the extent that holistic processing is believed to occur at perceptual as opposed to higher cognitive levels therefore, the dominant view is that it is special to faces.

1.4. Models and neuroanatomy of face processing

Thus far we have considered classic evidence for the exceptional processing of faces, and reviewed debates over the nature of this encoding strategy. How is this encoding strategy implemented? Bruce & Young's (1986) cognitive model was an early attempt to answer this, and sought to account for dissociations of face processing impairment in prosopagnosics. They noted both the occurrence of poor unfamiliar face processing combined with intact familiar recognition, and vice-versa. It is also possible to dissociate deficits of person memory from more specific face deficits (which allow continued recognition via non-face means; Young, 2011). These suggest a division of person recognition into separable face encoding, identity, and semantic information retrieval stages. Furthermore, patients' performance with facial identity is often dissociable from that of expression (Tranel, Damasio, & Damasio, 1988; Young, McWeeny, Hay, & Ellis, 1986), similarly suggesting separate processing of these attributes. Bruce and

Young's model posits an initial structural encoding stage, followed by separate routes for identity and expression processing. The identity route includes face recognition units and person identity nodes, whereas the expression route shares a view-centred encoding with facial speech analysis and directed visual processing (Figure 1.9).

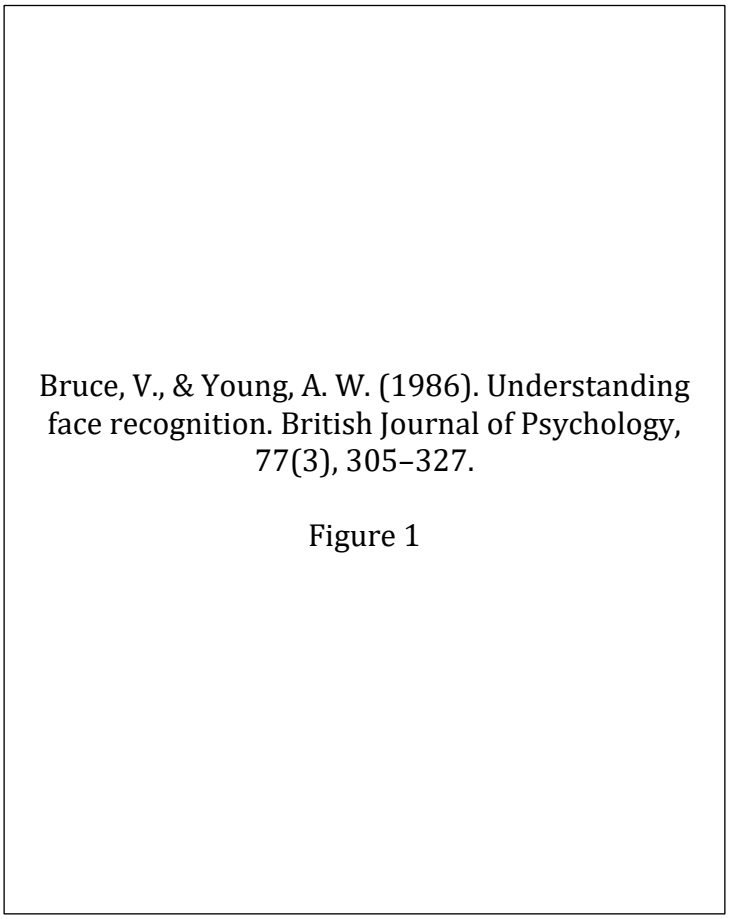


Figure 1.9. Bruce & Young's (1986) original cognitive model of face perception.

Common across the latter functions is a dependence on the changeable aspects of a face, in contrast with identity which is stable across expressions. Haxby, Hoffman, & Gobbini (2000) proposed a complementary neural model of the distributed face processing system, in which changeable versus invariant aspects of a face are processed by the superior temporal sulcus (hereafter STS) and FFA, respectively. This core system (including also the earlier inferior occipital gyri) then serves a multi-modal and multi-purpose extended system for attentional, emotional and semantic/biographical roles. With minor modification, Haxby et al's model accommodates changeable aspects as part of identity recognition (characteristic mannerisms), and conversely the inclusion of visual form in judgements of expression (O'Toole, Roark, & Abdi, 2002; Figure 1.10). Calder & Young (2005; Calder, 2011) go further in emphasising these crossovers, with a model that presents identity and expression as common visual processes, only segregated post-visually in the brain.

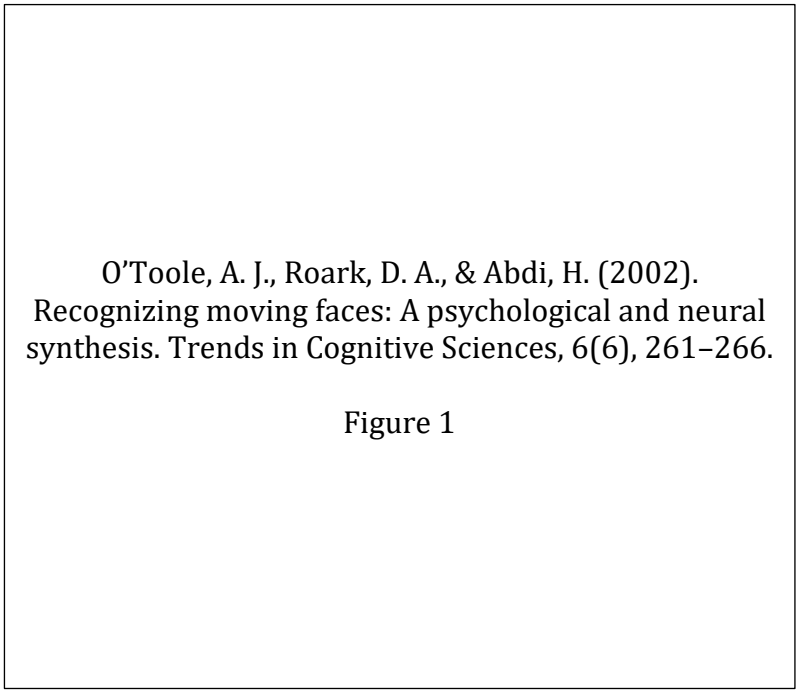


Figure 1.10. O'Toole et al.'s (2002) neurocognitive model of face perception, a revision of Haxby et al. (2000).

The inferior temporal gyrus receives input from early visual areas, and houses the occipital face area (hereafter OFA), the first stage of the core system. It responds preferentially to faces (Gauthier et al., 2000; Pitcher, Walsh, & Duchaine, 2011) with activity in this area associated with an electrophysiological response 100ms after stimulus onset, comparatively early in cortical terms and prior to the N170 associated with subsequent areas (Liu, Harris, & Kanwisher, 2002). The OFA is believed to represent face parts, with transcranial magnetic stimulation (hereafter TMS) over it impairing judgments of facial features while sparing inter-feature configural judgements (i.e. distances, Pitcher, Walsh, Yovel, & Duchaine, 2007). This concords with the wider role of lateral occipital cortex as an intermediary between low-level visual properties and complete object representations downstream (Grill-Spector et al., 1998). Whereas FFA and face-selective STS show higher fMRI responses to upright over inverted faces, OFA does not (Yovel & Kanwisher, 2005). FFA and OFA are further contrasted in fMRI studies showing greater activation for configuration or whole face context in the former, versus feature-driven activation in the latter (Liu, Harris, & Kanwisher, 2010; Nichols, Betts, & Wilson, 2010).

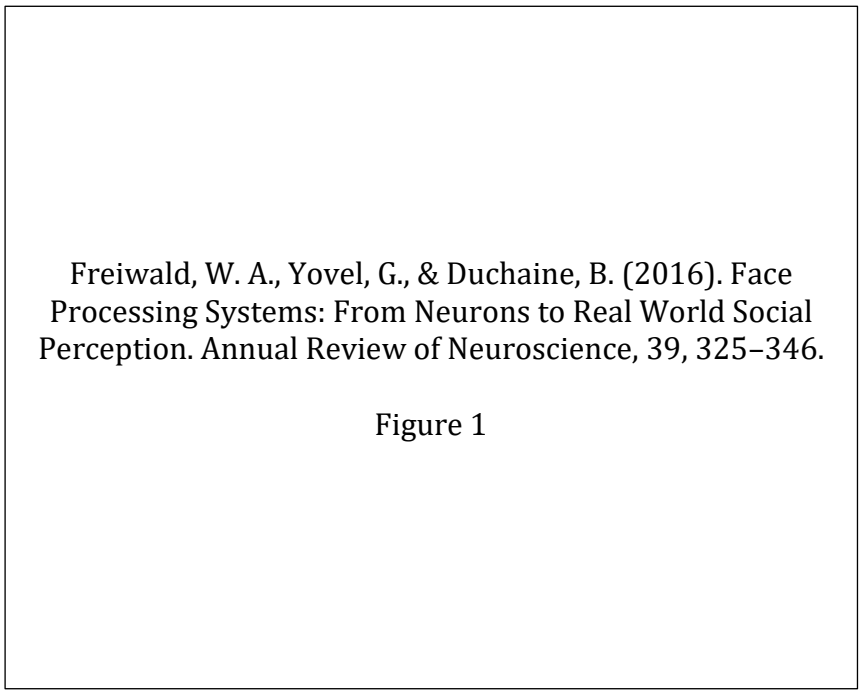


Figure 1.11. Schematic of the core face processing network plus anterior STS (see Chapter 5). From Freiwald, Yovel & Duchaine (2016).

Thus it is plausible that part-based representations in OFA are bound into whole face representations by subsequent areas, particularly FFA (Zhang, Li, Song, & Liu, 2012). Overall, models of the neural architecture subserving face processing share a division of labour reminiscent of the broader dorsal and ventral visual streams (Lehky & Sereno, 2007; Ungerleider & Mishkin, 1982), and a progression from simple components to complex abstractions (Figure 1.11).

1.5. Representing faces: spaces, norms and exemplars

In Marr's (1982) terms, we have thus far considered the computational (holistic processing) and implementation (cortical architecture) aspects of face perception, but not the algorithmic level. What representations and processes do these brain regions employ to achieve holistic face perception?

Generally, recognition from visual input is an ill-posed inverse problem (Troje, 2013). The utility of perception lies in representing what is in the world rather than what the world looks like per se (Frisby & Stone, 2010, p. 16; but see Purves, Monson, Sundararajan, & Wojtach, 2014), and this distinction is significant because otherwise irrelevant context causes identical sources to produce widely varying sensations. Stability of a representation under appearance-altering transformations such as lighting or viewpoint is referred to as invariance (DiCarlo, Zoccolan, & Rust, 2012).

Faces are a similarly appearing class of visual objects, but their social salience renders the need to discriminate between them especially acute. Invariance in this light may be cast as a categorisation problem, so an appropriate representation becomes one that allows effective subdivision of visual input according to task. The production of such a representation has been conceptualised as manifold flattening (DiCarlo & Cox, 2007). Neural populations in the retina and downstream implement high-dimensional spaces, with dimensionality equal to the number of neurons in the population and points in such spaces corresponding to a pattern of firing rates resulting from specific sensory input. A given view of one face will produce some pattern of responses in the retina describable as a point in 'retina-space', while all possible views of this face comprise a manifold in retina-space. The same applies for subsequent neural populations up the visual hierarchy.

One face can elicit a wide range of activity patterns in the retina dependent on viewing conditions, and a different face can nonetheless elicit similar activity patterns (e.g. by similar lighting and visual location). Distinguishability of the faces' identities is then hampered by similarity of the activity they elicit. In DiCarlo & Cox's terms, the two faces' manifolds in retina-space are entangled such that they cannot be cleanly separated by an appropriate hyperplane. DiCarlo & Cox therefore characterise the successive re-representation of objects and faces by neural populations along the ventral visual stream as being linked by transformations that flatten and align these manifolds for easier separability. In this way, greater invariance with which to perform effective input classification is achieved at higher levels of the visual system.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. Trends in Cognitive Sciences, 11(8), 333–341.

Figure 1

Figure 1.12. The task of categorising faces as belonging to different identities may be cast as hyperplane separation of manifolds. In their initial pixel representation, the manifolds of the two identities are not easily separable, but through successive flattening a neural representation may be arrived at in which it is possible. From DiCarlo & Cox (2007).

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

Figure 1

Figure 1.13. Stimulus face space constructed by Loffler et al. (2005). The space allowed generation of synthetic faces, and was based on parameterisation of basic geometric features of passport style photos.

Interest naturally turns to the properties of the spaces in which these manifolds reside. What features do their dimensions encode, and what is their topology? Differential BOLD responses to stimuli with varying degrees of naturalistic image structure (region-specific 'metamers') are beginning to make explicit the statistics to which these areas are sensitive (Movshon & Simoncelli, 2014). Many face spaces (Valentine, Lewis, & Hills, 2016) have been considered, and though their dimensions tend to originate from experimenter intuition more than neuroscience, both fMRI and electrophysiology have been applied to their study. Loffler, Yourganov, Wilkinson, & Wilson (2005) presented participants with simplified 'synthetic' faces retaining basic geometrical features from passport-style photographs (Figure 1.13), and found increasing BOLD response in FFA with increasing distinctiveness (distance from the mean feature set). They also dissociated neural populations encoding different identities (directions in their geometrically parameterised face space) using fMRI adaptation: synthetic faces varying in identity did not co-adapt, whereas those varying only in distinctiveness did.

Freiwald, Tsao, & Livingstone (2009) recorded single cell activity in macaque middle face region (roughly homologous with FFA though known to encompass several areas in a network, Moeller, Freiwald, & Tsao, 2008), while presenting 'cartoon' faces varying in pre-specified attributes such as inter-eye distance or face aspect ratio. They found ramp-shaped tuning curves for several of these dimensions, often extending out to extremes beyond biological plausibility. Individual cells tended to be tuned to few feature dimensions, and the authors noted the suitability of this scheme for emphasising difference from the average face in that overall population response variance would be highest for 'unusual' features.

Both emphasise distinctiveness from an average face, and this dominant norm-based coding view is well supported psychophysically (Rhodes & Leopold, 2011) though the primary alternative, exemplar-based coding, continues to find support (Ross, Deroche, & Palmeri, 2014; Storrs & Arnold, 2015). In introducing face spaces as a concept, Valentine (1991) found that faces independently rated as more distinctive were better recognised, implying again distance from some

central tendency (though he cautioned that this alone does not favour either type of underlying encoding). Conversely, Rhodes, Brennan, & Carey (1987) found that caricatures were identified more quickly than the veridical faces on which they were based. Their stimuli were line drawings traced from photographs, defined by 169 hand-placed key-points. The stimulus space was therefore essentially 169-dimensional, with the norm being the average of all faces' points and a caricature being a single face's points extended further from the norm but in the same direction. Evidence specifically for norm- over exemplar-based coding is stronger from adaptation studies such as Rhodes & Jeffery (2006), who demonstrated selective improvements in identification for faces lying opposite an adapter in face space. They used a morphing algorithm comprising dense laser-scanned meshes for 3D structure and equally numerous RGB values for texture of each face (Blanz & Vetter, 1999; Leopold, O'Toole, Vetter, & Blanz, 2001), amounting to a non-trivial and non-biologically derived stimulus space. It too allowed computing of a norm however, and therefore of 'anti-face' counterpoints to any point in the space by reflection through the origin (Figure 1.14). While adaptation to such anti-faces improves discriminability of the original (presumably by renormalising the space toward the former and away from the latter), no such improvement was found following adaptation to anti-faces not derived from the test face, but nonetheless matched in perceived dissimilarity.

In all cases, parameterisation of the stimulus faces (and therefore implicit dimensionality of the space) derives either from subjective ratings or the constraints and pragmatism of the experimenter. While such stimuli have been useful for probing perceptual face space, the possibility that they omit potentially relevant cues - resulting in theory beholden to design choices - remains (Dakin & Watt, 2009; Leder & Bruce, 1998). Principal component analysis allows the creation of an orthogonal space with minimal prior assumptions, based only on the variance of its input (Sirovich & Kirby, 1987; Turk & Pentland, 1991). Given that faces are similar in their appearance, it also provides an efficient encoding with 100×100 pixel grayscale images typically well represented by 50 principal components. Because of its unsupervised nature and grounding in image statistics, PCA frameworks have proven fruitful for modelling many aspects of

human face perception (Calder & Young, 2005). Notably, Calder & Young found PCA capable of encoding expression and identity within a single representation, challenging the dominant functional dissociation account.

To conclude, due to the high and unspecified dimensionality of faces innumerable representational spaces are possible, and many have been realised across experimental parameterisations. A distinction therefore needs to be

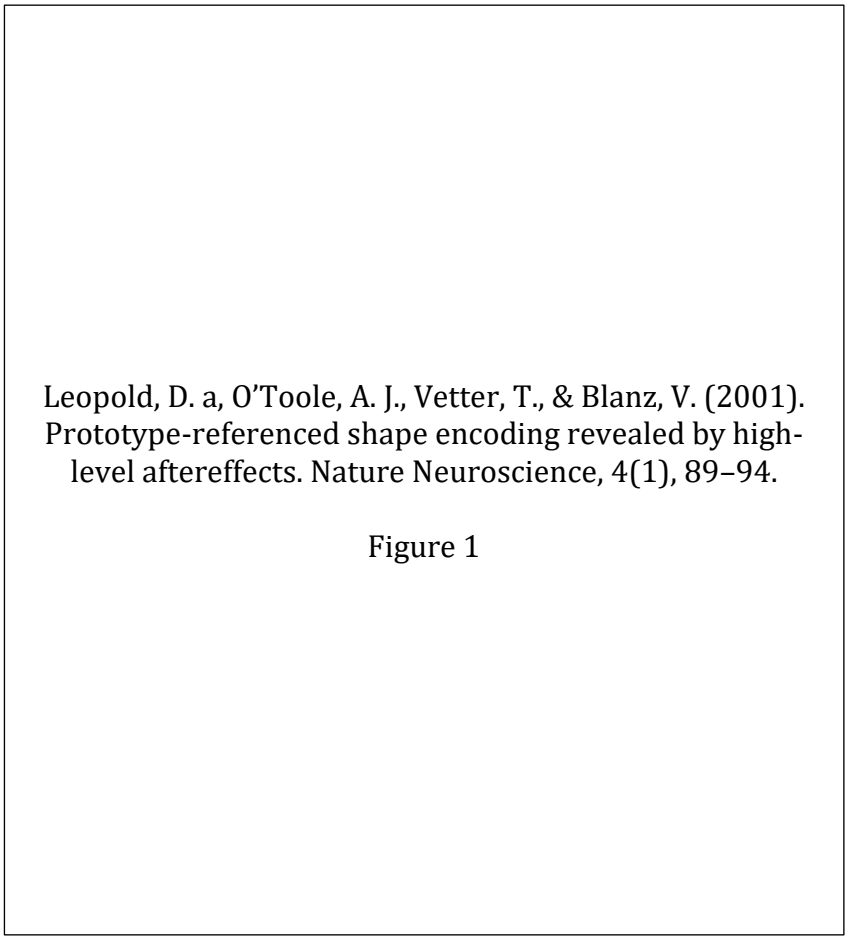


Figure 1.14. Stimulus face space constructed by Leopold et al. (2001). A face model was trained on a database of 3D face scans and textures, such that arbitrary new face images or scans could be well represented. This parameterisation of faces allowed calculation of an average face (origin), and 'anti-faces' by reflection through the origin.

drawn between stimulus face spaces, and those claiming to model perceptual processing. Regarding the latter, psychophysical and increasingly neuroscientific evidence (e.g. Chang & Tsao, 2017) favour a norm-based neural encoding over an exemplar-based scheme.

1.6. Expressions and dynamic face perception

Holistic face representations are likely to be informed by motion. A priori, it is not clear how an invariant representation could be generated without exposure to the possible variability (Johnston, 2011), especially in the case of non-rigid shape changes as induced by expressions. In other words, it is reasonable to suggest that stability of identity depends on prior knowledge of what may and may not change as a face emotes. The shape of the eyes may do so, but if the distance between them alters the image is probably of a different face.

In spite of this, face research was slow to shift from the dominant use of static stimuli and initial results were equivocal (O'Toole & Roark, 2011). Several studies failed to find an advantage for recognition of moving faces over static stimuli or could not be replicated (O'Toole, Roark, & Abdi, 2002 provide an overview). It seems that static information is sufficient for high performance, so that revealing any motion advantage requires subtraction of this information while leaving motion intact.

Logically one might question the extent to which motion cues in stimuli can be preserved while subtracting the form carrying that motion. Hill & Johnston (2001) used a 3D model averaged from the shapes of 200 individuals, animated by marker tracking of the facial actions of 12 volunteers' faces as they told a joke (Figure 1.15). Observers were able to identify individuals and their gender based on this animation, despite the use of a common model. Lander & Bruce (2000) degraded videos of famous people using binary thresholding (Figure 1.16) and compared recognition performance for frame-jumbled, sped up and reversed sequences against the originals. Performance was better for the originals in all cases except reversed (which just failed to reach significance). These and similar findings led O'Toole, Roark, & Abdi (2002) in their revision of Haxby et al's model to posit the involvement of motion in recognition only when viewing conditions are poor. They also argued that familiarity was required before identification based on an individual's characteristic mannerisms (dynamic identity signatures) was possible. This forms the supplemental information hypothesis. Evidence for an alternative representation enhancement hypothesis, that motion benefits identification by providing greater opportunity to learn the

3D structure of the face (structure from motion) is less well supported (O'Toole & Roark, 2011).

Hill, H. C. H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. Current Biology, 11(11), 880–885.

Figure 1

Figure 1.15. 3D model employed by Hill & Johnston (2001) and the marker tracking that drove its motion. Note that the appearance of the model remained the same across actors, so the stimuli only differed by motion content.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. Ecological Psychology, 12(4), 259–272.

Figure 2

Figure 1.16. Example binary-thresholded stimuli used by Lander & Bruce (2000).

Despite being inherently dynamic, expression perception research is also dominated by use of static stimuli (Cunningham & Wallraven, 2009). This partly follows from an emphasis on Ekman's (Ekman, Sorenson, & Friesen, 1969) basic expressions, of which there may be six, seven (Ekman & Heider, 1988), or four (Jack, Garrod, & Schyns, 2014). These basic expressions are held to be fundamental based on their universality across cultures, and may derive from biologically adaptive behaviours (such as avoidance of noxious stimuli by

wrinkling the nose) that have been co-opted for social signalling (e.g. disgust; Chapman, Kim, Susskind, & Anderson, 2009). They are also straightforward to represent via their maximal 'apex' posture however, so is perhaps not surprising that viewing facial motion in a ramp from neutral to apex does not improve performance over static presentation of the final frame alone (Fiorentini & Viviani, 2011) – these stimuli fail to uncover the motion advantage.

Meaningful interpretation of expressive dynamics using Ekman's basic six expressions has been achieved, however. Schyns and colleagues (Jack & Schyns, 2015) combine the Facial Action Coding System (hereafter FACS; another stimulus face space which segregates movement into discrete action units, hereafter AUs, derived from the musculature of the face, Ekman & Friesen, 1976a; Figure 1.17) with information theory, EEG and behavioural methods to decompose the unfolding motion signal. By presenting randomised AUs and reverse-correlating participants' categorisations of the apparent emotion, Yu, Garrod, & Schyns (2012) developed a Generative Face Grammar mapping these emotions to their constituent AUs and temporal dynamics. Jack, Garrod, & Schyns (2014) then computed the probability of AUs peaking across the sequence in terms of Shannon entropy, revealing that biologically relevant actions (e.g. nose wrinkling) tend to occur early. An ideal Bayesian classifier further revealed how later actions serve to clarify the initial signal, with early confusion between surprise/fear and disgust/anger resolving over the course of the sequence.

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10), 974–989.

Figure 2

Figure 1.17. Some example facial action units, with their difference (subtraction) from the neutral posture. AUs are dynamic and follow a motion trajectory from neutral to apex posture. From Donato, Bartlett, Hager, Ekman, & Sejnowski (1999).

Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of Visual Information Integration in the Brain for Categorizing Facial Expressions. Current Biology, 17, 1580–1585.

Figure 1

Figure 1.18. Bubble occlusion employed by Schyns et al. (2007). Stimuli were decomposed into frequency bands to be sampled separately by Gaussian apertures matched in cycles per face.

Using bubble occlusion (Gosselin & Schyns, 2001) and EEG, Schyns, Petro, & Smith (2007) probed how features are integrated prior to the N170 component. By obscuring their facial stimuli save for randomised windows permitting specific spatial frequencies, they were able to link behavioural categorisations of the basic emotions to concurrently recorded activity (Figure 1.18). They found evidence of feature integration over time, tending to proceed from the eyes downward and stopping once diagnosticity had been achieved.

Expressions in the wild are typically more nuanced, but these conversational expressions have received comparatively less attention. A conversational partner is much more likely to express agreement, thinking or confusion than anger or disgust (Cunningham & Wallraven, 2009), yet the subtlety of these signals renders their interrogation difficult (Bülthoff, Cunningham, & Wallraven, 2011). Often they serve as back-channel communication augmenting the verbal stream, prompting the speaker to for example continue or re-explain (Yngve, 1970, as cited in Bülthoff, Cunningham, & Wallraven, 2011). Even where basic expressions are concerned, paradigms involving simple classification neglect the subtle modes of variation they may take – temporal structure can make a smile appear friendly or malevolent, for example (Johnston, 2011). A related issue is the perceived genuineness of an expression (Krumhuber, Manstead, & Kappas, 2007), which is especially pertinent given the volume of research based on stimuli mimed by actors.

In a complementary approach to Schyns and colleagues' reverse correlation of basic expressions to AUs, it is possible to decompose naturalistic facial motion sequences using PCA. Additional information representing the deformation of the face within each frame is required however, to achieve smooth gradation of posture over the dimensions of the space. This is because images (or frames of a sequence) become represented as linear sums, so their back-projected reconstruction is essentially a weighted average. This leads to blurring or 'ghosting' where contours of the images representing principal components do not correspond; though this may be tolerable for static images it does not lead to coherent motion in the dynamic case.

Berisha, Johnston, & McOwan (2010) achieve this by describing each frame's posture relative to a common reference. They do so using a biologically motivated optic flow algorithm (Johnston, McOwan, & Buxton, 1992) to produce vector fields warping each frame to the reference. These dense vector fields then form part of each frame's vector description (alongside texture) for submission to PCA. Berisha and colleagues examined how bubble occlusion would affect subsequent reconstruction fidelity, and found the areas around the eyes and mouth to be particularly important (Figure 1.19).

Berisha, F., Johnston, A., & McOwan, P. W. (2010). Identifying regions that carry the best information about global facial configurations. Journal of Vision, 10(11), 27–27.

Figure 6

Figure 1.19. Regions that carry the best information about global facial configuration. Berisha et al. (2010) constructed a PCA face space from sequences of facial motion, and projected occluded sequences into it to assess which regions were most important for accurate reconstruction.

As per Haxby et al. and O'Toole & Roark's models, facial dynamics are believed to be processed separately from invariant features in the STS. This area is less studied than the FFA, and there is more disagreement over its connectivity and functional specificity. It has been suggested that the STS in fact lies outside the OFA-FFA network, with distinct ventral white-matter connections (Gschwind, Pourtois, Schwartz, Van De Ville, & Vuilleumier, 2012, see also Chapter 5 below), and many fields beyond face perception have claimed it as 'their area' (Hein & Knight, 2008). This notwithstanding, STS is known to respond more to faces than objects, and particularly to moving faces (Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011). Furthermore, the particular preference of STS for both facial and bodily motion (reviewed by de Gelder et al., 2010) is starting to become elucidated. Alongside a broader trend for the subdivision of these once monolithic areas into smaller interconnected patches (Tsao, Moeller, & Freiwald, 2008), Weiner & Grill-Spector (2013) in their review of higher-level visual

cortical organisation describe alternating face- and body-selective cortex bordering area MT (Figure 1.20). Their appraisal of these distributed patches leads them to propose a new three-stream model of visual processing, with the inclusion of a multimodal lateral route.

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. Psychological Research, 77(1), 74–97.

Figure 10

Figure 1.20. Face and limb selective regions neigbour each other, radiating around human middle temporal complex and extending ventrally. MT = middle temporal area; MST = medial superior temporal area; ITG = inferior temporal gyrus; IOG = inferior occipital gyrus; LOS = lateral occipital sulcus; pSTS = posterior superior temporal sulcus; MTG = middle temporal gyrus; pFUS = posterior fusiform gyrus; OTS = occipitotemporal sulcus; mFus = mid-fusiform sulcus. From Weiner & Grill-Spector (2013).

In conclusion, though temporal dynamics play a logically central role in face perception theory (such as development of invariance), research was slow to find a role for motion over and above performance achievable with static faces. Mirroring the dependency of face spaces' characterisation on experimenter design choice, this proved largely to be a question of developing appropriate stimuli, hindered at least initially by the added complexities of the time domain and its dissociation from facial appearance.

Facial expressions entail a tremendous degree of visual and conceptual nuance, of which only a small, tractable corner has yet been studied in detail. The

temporal dynamics of basic expressions have been shown to transmit evolving information over time as an initially coarse emotional signal is refined, while randomised spatial occlusion suggests observers may parse facial expressions from the top downward. These – and similar applications of occlusion with an information-theoretic approach to more naturalistic facial movement – implicate the eye and mouth areas as key to the fidelity of the expressive signal.

The STS is the primary dynamic processing area of the core face network, though its functions are many and it is less well characterised than the FFA. Recent neuroscience casts doubt on shared cortical origins for the two face streams beyond V1 (or indeed there being two streams), and it is becoming increasingly clear that once monolithic cortical areas actually subdivide into smaller interconnected patches.

1.7. Temporal relationships and cross-feature interactions

Recently, Cook, Aichelburg, & Johnston (2015) reported evidence for global perceptual models of facial motion. Akin to the composite face effect with static stimuli, they found interactions of perceived motion across facial features that occurred for upright, but not inverted faces.

Cook et al. presented participants with pairs of animated test faces either side of fixation, whose eyes and mouths moved periodically. Although the stimuli were not naturalistic videos therefore, the stimuli provide the advantage of complete control over and reproducibility of the head models' facial motion. The mouths opened and closed sinusoidally while the eyes opened and closed linearly (eyelid movement thereby followed an imperfect square wave with constant rise/fall times). The eyes and mouth moved at the same frequency, but the rise and fall times of the comparison face's eyes was varied so as to allow fitting of psychometric functions and estimation of points of subjective equality (hereafter PSEs).

Cook et al. presented participants with pairs of faces upright or inverted, and asked them which of the two had the faster eye blinks. They also varied the timing of the standard's mouth movement relative to the eyes, so that it could could lead, trail, match or oppose (Figure 1.21). They found slower PSEs for upright faces relative to inverted faces only when mouth movement matched or

led the eye blinks – in other words, an orientation-dependent illusory slowing of perceived eye blink speed that was contingent on relative timing of the mouth. A second experiment discounted possible origins in motion profile (sinusoidal versus constant velocities), by swapping them yet reproducing the inversion

Cook, R., Aichelburg, C., & Johnston, A. (2015). Illusory Feature Slowing: Evidence for Perceptual Models of Global Facial Change. Psychological Science, 26(4), 512–517.

Figure 1

Figure 1.21. Schematic of Cook et al.'s (2015) standard face motion. The eyes and mouth could move at one of four phase relationships, and the speed at which the eye blinks transitioned from closed to opened or vice-versa was varied.

effect of leading mouth movement on perceived eye blink speed.

These results are notable for several reasons. Primarily, they represent an extension of the substantial literature on holistic binding and cross-feature interactions to the dynamic face domain, showing that movement of one feature can interfere with perception of the movement of another even when not relevant to the task and at the expense of veridical perception. The presence of an inversion effect argues against lower-level generic visual mechanisms in favour of face-specific processing. Importantly, the dependence of the effect on relative feature timing highlights the intrinsically temporal nature of this binding.

Analogous to the holistic eigenfaces produced by PCA on static images, PCA on dynamic sequences as described in previous sections results in components encoding coordinated motion of several facial features. Behavioural results with static stimuli have emphasised holistic processing as a likely consequence of

learning invariance, to allow efficient representation of faces by characterising how they differ. Cook et al's findings suggest an equivalent process for facial movement, whereby correlations are exploited and perceptual priors formed that encode overall dynamics.

The following chapters detail several investigations into the role of relative timing on the binding of moving facial features. Using psychophysics, adaptation, and fMRI, we extend Cook et al.'s initial result and characterise the role that temporal coincidence plays as a key component of the dynamic visual stream of facial expression. Our first aim was to develop a psychophysical task and stimuli that could measure sensitivity to feature motion while allowing principled manipulation of the wider dynamic facial context. If the perceptual bias reported by Cook et al. does indeed reflect cross-feature interdependencies, it is reasonable to hypothesise that these would manifest also in variable perceptual sensitivity (see Chapter 2 for the theoretical advantages of measuring sensitivity over bias). It was also important to verify that the new paradigm measured face perception specifically, as opposed to generic or lower-level visual processes.

Our second aim was to apply this paradigm to facial timing, seeking to both replicate Cook et al. and reconcile their findings with related literature. We hypothesised that the cross-feature interactions measured by our paradigm would indeed be sensitive to such elementary temporal structure, though (as detailed in Chapter 3) which specific timings are relevant was hitherto a matter of some disagreement.

The remainder of the present work aimed to place our findings in a neuroscientific context, assessing cerebral lateralisation, visual adaptation and finally cortical activation as measured via differential BOLD responsivity. The extant face perception literature and Cook et al. specifically suggest hypotheses regarding the first two, but (as will be seen in Chapters 3 and 4) our behavioural results discounted these in favour of alternative explanations. Chapter 5 assessed the final hypothesis that facial motion sensitive brain areas are modulated by feature timing using a simple whole-brain analysis.

CHAPTER 2. Dynamic features are not processed independently

2.1. Introduction

The chapter begins with a general overview of those aspects of the methods that were common to most of the experiments described, so as to avoid repetition in subsequent per-experiment methods sections. The key findings to be reported in this section are outlined below.

Experiment 1 found an illusory slowing of facial movement. In this case the perceived slowing applied globally to the face stimuli and resulted from facial motion adaptation rather than the effect of one feature acting upon another as described in Cook et al. (2015). This result served to confirm that our stimuli – particularly the replacement of eye blinks by eyebrow movement –remained an effective tool for investigating the holistic binding of dynamic faces.

Returning to cross-feature interference, Experiment 2 examined whether movement in one feature can reduce sensitivity to the movement of another. An effect of mouth movement on discriminability of eyebrow movement was found that was specific to upright faces.

Experiment 3 was an attempted replication of Experiment 2 with several amendments to the design aimed at strengthening the effect. However, no cross-feature interactions for either face orientation was found in this case, which we attribute (referencing previous literature) to these new features of the design.

2.2. General Methods

Almost all of the experiments had the shared property of manipulating movement in one stimulus feature (mouth), while measuring its influence on perception of another moving feature (eyebrows). Given the information required to complete the tasks was localised to the eyebrow region, changes in performance induced by manipulating the mouth provide evidence for cross-feature interactions in processing of facial movement.

To allow precise control of this movement, we developed stimuli using a computer-generated head model whose eyebrow and mouth vertical positions were parameterised (Figure 2.1). Trials in our experiments entailed presenting pairs of these animated avatars for a forced-choice judgement (or matching by adjustment in Experiment 7). Eyebrow and mouth movement oscillated sinusoidally: raising/lowering or opening/closing respectively. Except in Experiments 1 and 7, participants had to indicate which of the two faces' eyebrows were misaligned (i.e. oscillating asynchronously so as to appear 'Mexican wave'-like; see supplemental video). The majority of our experiments therefore measured discriminability of this unusual feature motion.

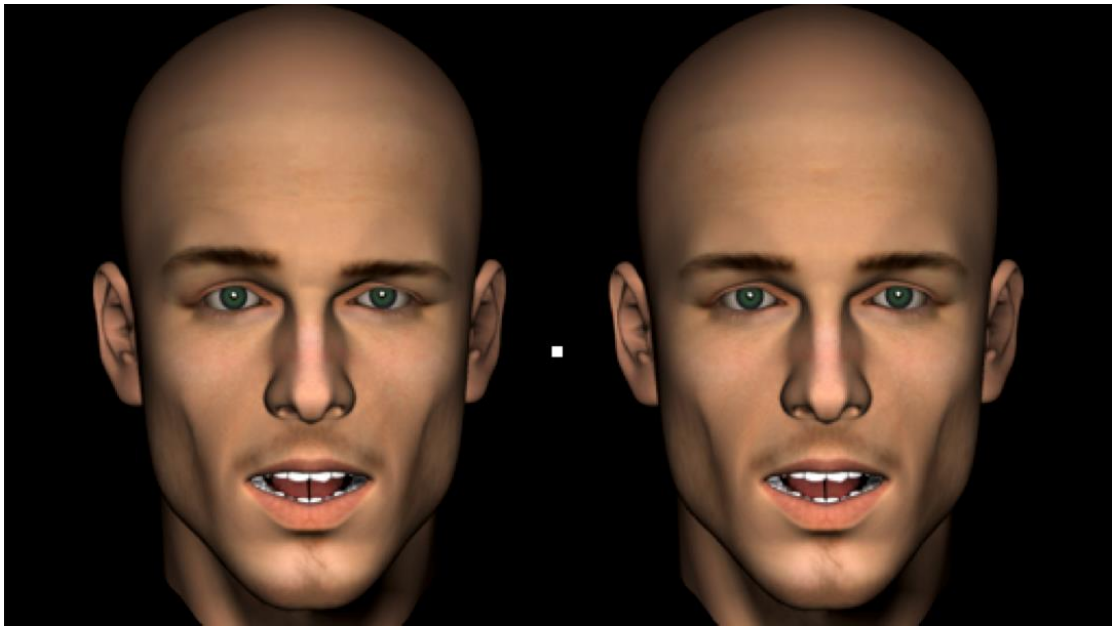


Figure 2.1. An exemplar frame illustrating the face model used throughout. Vertical positions of eyebrows and mouth are independently controllable, allowing their oscillatory animation.

Experiment	Methodology	Dependent variable	Eye tracker	Inversion	Test stimulus presentation time	Inter-stimulus interval	Feedback	Progress counter	Total trials	Total time (approximate)
1	2AFC bias (following adaptation)	PSE (face rate)	✗	✓	2s	3s	✗	✗	280	1hr
2	2AFC sensitivity	Threshold (eyebrow misalignment)	✗	✓	3s	2s	✓	✗	1120	2hr
3	2AFC sensitivity	Threshold (eyebrow misalignment)	✓	✓	3s	2s	✓	✓	1120	2hr 45m
4	2AFC sensitivity	Percent correct (eyebrow misalignment)	✓	✓	3s	1s	✓	✓	720	2hr
5	2IFC sensitivity	Percent correct (eyebrow misalignment)	✓	✓	3s	1s	✓	✓	640	3hr
6	2AFC sensitivity (following adaptation)	Percent correct (eyebrow misalignment)	✓	✗	3s	1s	✓	✓	800	3hr 15m
7	Method of adjustment (following adaptation)	PSE (eyebrow-mouth phase)	✗	✗	2s	1s	✗	✓	800	1hr – 2hr 15m

Table 2.1. Some common parameters listed for each psychophysical experiment.

The remainder of this section provides a general overview of our approach, many aspects of which apply across the seven experiments. When introducing each one, subsequent sections then need only focus on deviations from the overall scheme. Some typical parameters are further listed in Table 2.1.

Participants

All included participants reported normal or corrected-to-normal vision (note the detail on eye tracking in Design and Procedure below). They were recruited either opportunistically or via university research participation schemes for monetary remuneration. No special populations were tested and no participant reported diagnoses of visual or psychological impairment. Some participants (including both authors) participated in multiple experiments, and so will have had greater familiarity with the task and stimuli. All comparisons were within-participant, however.

Stimuli and Apparatus

The eyebrows and mouth of a male 3D head model were 'morphed' using Smith Micro Software's Poser 9. In this context morphing refers to a smooth transition from one posture to another. Poser predefines many trajectories along which model features may vary, including the vertical position of the brow or jaw. The head faced forward (in portrait) with a single, even light source positioned behind the camera. The face was therefore neither lit from above nor below, avoiding a potential confound with this cue during stimulus inversion. Of Poser's available models, "James" was chosen for its lack of substantive facial or head hair, lifelike skin texture (including non-rigid deformability) and well-defined structural contours (e.g. cheekbones).

Frames of a single face executing a half-cycle of both eyebrow and mouth were exported from Poser, to be arranged into multi-cycle sequences of two faces immediately prior to presentation (Figure 2.2). Over the 5040 exported frames (hereafter 'basis frames') mouth motion ranged from closed to maximally open, while the eyebrows morphed through lowered to maximally raised. Animation was achieved by sequential presentation in alternating forward and reversed order, while rate was determined by subsampling: using only every 200th basis frame results in features oscillating at around 1.2Hz on a 60Hz monitor for example. Rate only varied in Experiment 1; all faces oscillated at 1.5Hz across Experiments 2 – 7.

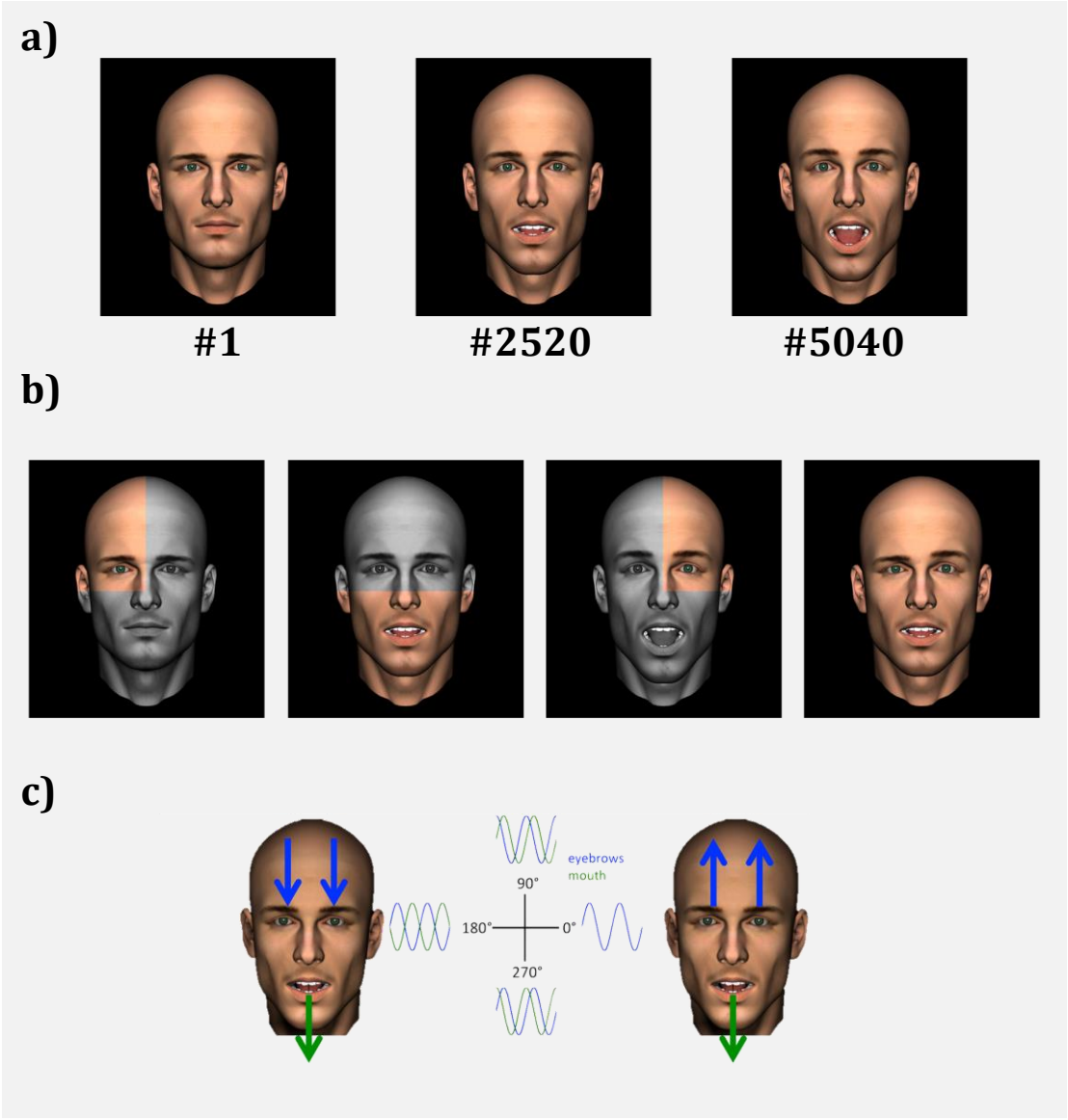


Figure 2.2. a) All presented stimulus frames are constructed from a subset of 5040 basis frames generated by Poser 9. b) Each stimulus frame is constructed by stitching together eyes and mouths of different basis frames. c) Following Cook et al.'s convention, peaks of the features' driving sinusoids correspond to mouth open or eyebrows raised.

Phase offsets between the mouth and individual eyebrows were achieved by 'stitching together' bottom halves and top quarters of different basis frames as required. We define 0 degrees of relative phase between eyebrows and mouth as the mouth being maximally open while the eyebrows are maximally raised. 180 degrees of feature relative phase therefore denotes eyebrows lowering while mouth opens, 0 – 180° denotes eyebrows raising before mouth opens, and 180 – 360° denotes mouth opening before eyebrows raising (Figure 2.2).

The head's initial resolution was 227×333 pixels it but was rescaled to subtend 8 degrees of visual angle (hereafter DVA) for Experiment 1, or 5 DVA for Experiments 2 – 7.

Two such faces (standard and comparison) were presented within each trial. In all bar Experiment 5, they appeared simultaneously either side of and near to fixation. In Experiment 5 the faces appeared sequentially in just one of these positions such that presentation was confined to a single hemifield. Because each basis frame included a portion of neck underneath (sufficient space to accommodate the chin while mouth was maximally open), the fixation point was shifted downward by 12.5% of image height so as to align it with the vertical midpoint of the nasal ridges (Figure 2.1). Fixation was likewise shifted upward when the faces were inverted. Stimuli were presented against a black background in all bar Experiment 7, wherein the background colour was the average of the faces' in Lab colorspace (so as to match the background colour to the faces' skin tone).

Our experiments were implemented using custom MATLAB code and Psychtoolbox-3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Stimuli were presented in a darkened room, on a Mitsubishi Diamond CRT monitor. Participants responded using the computer keypad, or a Griffin Powermate rotatable dial in Experiment 7. Experiments 3 – 5 employed a Cambridge Research Systems HS-VET Mk. 2 eye tracker. The Snellen chart used prior to the eye-tracked experiments was downloaded from www.provisu.ch.

Design and Procedure

Participants used a chin-rest and fixated on a central point, thus controlling the area of retina stimulated (there was no free viewing).

Except for Experiment 7, trials consisted of presenting test stimuli for two or three seconds followed by a response interval in which participants made their forced choice. This response interval lasted a minimum of 1 – 3 seconds (details in Table 2.1 and below) with no maximum time limit, although participants typically responded within one second. The response interval included text instructing the participant to respond, and once they had done so feedback ("[in]correct" coloured text in Experiments 2 – 6) and/or a trial counter signalling progress through the current block (in Experiments 3 – 7) was displayed for the remainder of the interval. Feedback and trial counters were intended to maintain participants' motivation throughout.

Experiments 1 – 5 included an orientation independent variable, with the test stimuli inverted in half of the blocks (ordering was randomised). This served to test whether any effects found were ascribable to face-specific processing mechanisms (see Introduction). With face specificity thereby demonstrated, Experiments 6 and 7 presented upright stimuli only.

Experiments 1 – 3 employed the method of constant stimuli to estimate psychometric functions for a point of subjective equality or threshold. Experiments 4 – 6 measured percentages correct at a single stimulus level, fixed at the approximate 75% threshold ascertained from Experiment 2. Experiment 7 employed the method of adjustment to acquire repeated estimates of a standard over time. Further details of these independent variables and other experiment-specific dependent variables are provided in those experiments' methods sections below.

In Experiments 1 and 6, each trial was preceded by a period of adaptation – 30 seconds for the first trial of a block and 5 seconds subsequently. Experiment 7 aimed to measure the build-up of adaptation using method of adjustment rather than forced choice, therefore its design differed substantially and is detailed in 4.3.2 below.

Experiments 3 – 6 utilised an eye tracker to enforce fixation. Because the eye tracker could not accommodate spectacles, participants in these experiments either wore contact lenses or were uncorrected. To ensure sufficient acuity in either case a brief Snellen test was used for initial screening, and anybody unable to achieve 20/20 performance was excluded. The eye tracker necessitated an additional calibration stage prior to each block, and drift correction prior to each trial. During drift correction, a pair of static face images with mouths closed and eyebrows aligned occupied the stimulus locations. Because drift correction could take a variable amount of time (typically around 600 – 1500ms) a 'countdown' whereby fixation colour changed through red, green and white in 200ms increments signalled progress toward stimulus onset as a cue for participants.

Enforcement of fixation entailed checking that participants' gaze remained within a small rectangular region around the fixation point, and rejecting the trial upon lapses. In these cases, ISI text informed of the lapse rather than requested a response, and the trial was appended to the end of the block for repeating (thus blocks lengthened with insufficient fixation discipline, providing enough motivation to keep typical lapse rates under 10%). Some participants' eyes were not well suited to eye tracking leading to a high number of spurious lapses, and poor calibration and drift correction. These participants were excluded.

2.3. Experiment 1: Adaptation to fast facial motion causes illusory slowing

2.3.1. Introduction

The first experiment served to pilot the new stimuli. Although based on the experiments described in Cook et al. (2015), it was necessary to ensure an effective means of probing dynamic face perception was retained after several adjustments to the stimuli. Whereas Cook et al. paired sinusoidal mouth movement with ramped eyelid movement, we replaced the latter with sinusoidal eyebrow movement. This was motivated by the greater amount of information carried by the eyebrows about overall facial posture (Berisha et al., 2010), the importance of their movement as components of expressive signalling (Jack et al., 2014), and their tendency to naturally covary with mouth movement (Berisha & Johnston, unpublished data). Driving both features' movement using the same elementary functional form is also arguably a desirable simplification.

We also sought to extend the scope of the facial movement speed illusions described by Cook et al. Whereas they employed global facial movement at a fixed rate of 1.25Hz and found their speed illusion in a specific component of one feature's motion (eye-blink transition), we asked whether global facial movement as a whole was similarly susceptible to illusory slowing.

We therefore asked participants to simply compare the overall relative speed of faces, and employed adaptation to attempt to manipulate this judgement. Although "adaptation to faces is among the most widely studied classes of high-level aftereffects" (Webster, 2015), adaptation to facial motion is less studied (exceptions include Curio, Giese, Breidt, Kleiner, & Bülthoff, 2008; de la Rosa, Giese, Bülthoff, & Curio, 2013; Laurence, Hole, & Hills, 2014). Like Cook et al., we used inversion to test whether any perceptual bias induced by adaptation could be attributable to face-specific mechanisms. We therefore hypothesised firstly that adaptation to fast facial motion would make subsequent faces appear slower, and secondly that this illusory slowing would occur for upright, but not inverted faces.

2.3.2. Methods

Participants

Data from 21 participants was collected, but one was excluded due to very shallow and poorly fitting psychometric functions, indicative of an inability to complete the task. Of the remainder, 8 were male and 12 female. Age ranged from 18 – 43 years (mean 25). Participants were recruited through an advert to the UCL Subject Pool and paid £6, with one exception who was known to the experimenter.

Stimuli and Apparatus

In each trial following adaptation, both faces were presented simultaneously. The faces were 8 DVA wide and were located 2 DVA apart. Stimuli were presented against a black background. Following responses, no feedback or trial counter was displayed. Stimuli were presented on a 20in monitor 1280 pixels wide by 1024 pixels high, with a refresh rate of 75Hz and at a viewing distance of 57cm. No eye tracker was used.

Design and Procedure

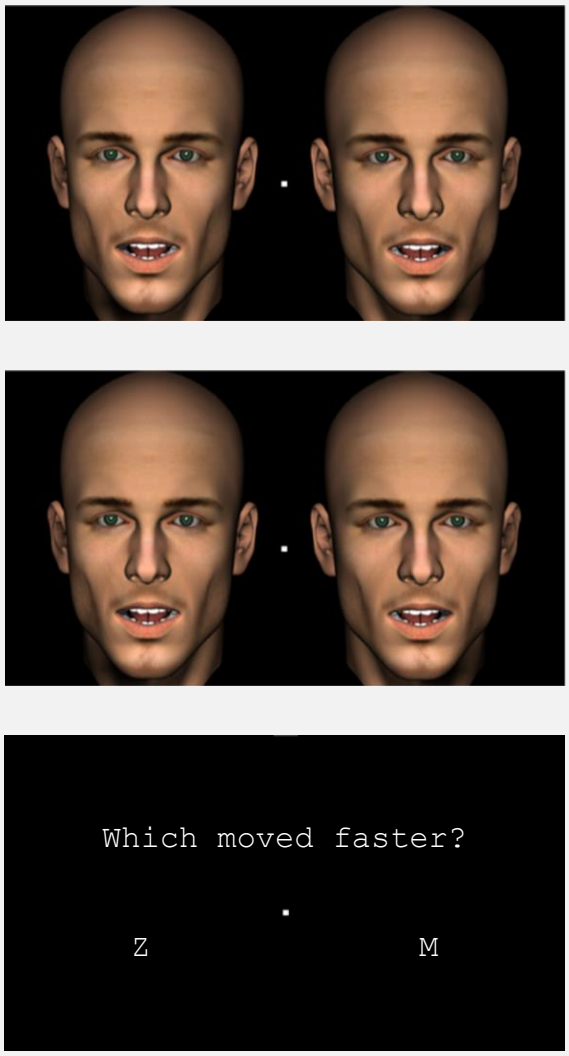
Participants completed a total of 280 trials split over 4 blocks (with brief rests in between), and the experiment took around one hour overall. Each block began with 30 seconds of adaptation in one of the two face positions (counterbalanced) while the other side contained no stimuli. The adapter oscillated at 3Hz as this is around the upper limit of high-level temporal binding (Holcombe, 2009; Nishida & Johnston, 2010). All face's eyebrows and mouths oscillated at 270 degrees of relative phase, such that the mouth was maximally open one quarter-turn before the eyebrows were maximally raised. This follows from the relative feature timing at which Cook et al.'s illusion was strongest. There was no misalignment of the eyebrow movement in any faces.

All faces were upright in half the blocks and inverted in the other half (counterbalanced). Following adaptation (initial or top-up

), the pair of test faces was presented for 2 seconds. The standard face always oscillated at 1.5Hz while the comparison face ranged from 0.75 – 2.25Hz in increments of 0.25Hz. We therefore took 10 measurements per level per orientation (10 × 7 × 2 × 2 = 280). Additionally, the two faces' starting phases were offset by a quarter turn: the standard initiated movement at 315 degrees of phase while the comparison began at 45 degrees of phase. This proved necessary following piloting to prevent participants using the divergence of feature alignment across the faces over time as a cue to their relative speeds.

Participants were then asked to indicate which face appeared faster in an inter-stimulus interval lasting at least 3 seconds (hereafter ISI; participants typically responded within 1 second). Although we were interested in measuring the effect of adaptation on the standard, it was necessary to randomise which test face occupied the adapted position so as to prevent participants realising one face's speed never changed (and subsequently basing their judgement solely on the comparison face in the non-adapted position). See Results below for details of the trials retained for analysis.

To summarise, we measured the proportion of times the comparison face was judged to be faster over a range of speeds relative to a fixed-speed standard occupying a region adapted to fast facial motion, and this was repeated for both upright and inverted faces.



Adaptation
30sec initial / 5sec topup
3Hz oscillation rate (both faces)
270° relative feature phase (both faces)
Upright or inverted (both faces)

Test
2sec
1.5Hz oscillation rate (standard face)
0.75 – 2.25Hz (0.25Hz steps; test face)
270° relative feature phase (both faces)
Upright or inverted (both faces)

Inter-stimulus interval
3sec +

Figure 2.3. Experiment 1 trial design. Participants adapted to the same fast-moving stimuli in both test face positions, followed by presentation of test faces that differed only in their rate of motion. Participants judged which test face was faster and did not receive feedback or notification of progress.

2.3.3. Results

Given the effect of adaptation on the comparison is not easily interpretable (one cannot assume the effect varies linearly across the standard – comparison relative speeds, for example), the half of trials where the comparison appeared on the adapted side were discarded. For each participant, psychometric functions for each of the two face orientations were calculated using Psignifit 3.0 (Frund, Haenel, & Wichmann, 2011), all of which fitted well (see Figure 2.4 for details). Points of subjective equality (hereafter PSEs) were then taken as the comparison speed at which the comparison was judged to be faster 50% of the time (as estimated by the midpoint of the fitted function).

Psychometric functions were computed separately for each participant and condition. The model to be fitted took the form

$$\Psi(x;\theta) = \gamma + (1-\gamma-\lambda)F(x;\alpha,\beta), \theta = (\alpha,\beta,\lambda,\gamma)$$

The function $F(x;\alpha,\beta)$ to be fitted was a cumulative Gaussian of the form

$$F(x;\alpha,\beta) = \Phi(g(x,\alpha,\beta))$$

where $g(x,\alpha,\beta)$ is the linear function

$$g(x,\alpha,\beta) = \frac{z_0}{\beta}(x-\alpha) + z_1$$

with z_0 and z_1 derived from the shape of $\Phi(x)$ such that β corresponds to the width over which $F(x;\alpha,\beta)$ rises from 0.1 to 0.9 after fitting. $\Phi(x)$ is the cumulative Gaussian function

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt$$

$\Psi(x;\theta)$ was fitted using constrained maximum likelihood estimation, with both the guess rate γ and lapse rate λ constrained to lie between 0 and 0.1 for experiment 1. For experiments 2 and 3, γ was constrained to 0.5. Goodness of fit was assessed by bootstrap generation of 2000 samples, with deviance of our observed sample from the fitted model considered acceptable if falling within the 95% confidence intervals of the bootstrapped distribution.

Figure 2.4. Estimation of psychometric functions for experiments 1, 2, and 3.

For upright faces, the mean PSE was significantly lower than veridical as revealed by one-sample t-test against 1.5Hz (M = 1.45Hz, SD = 0.08Hz; $t(19)$ = -2.82, two-tailed p = .01).

For inverted faces however, the mean PSE did not differ significantly from veridical (*M* = 1.49Hz, *SD* = 0.12Hz; *t*(19) = -0.46Hz, two-tailed *p* = .65; see Figure 2.5).

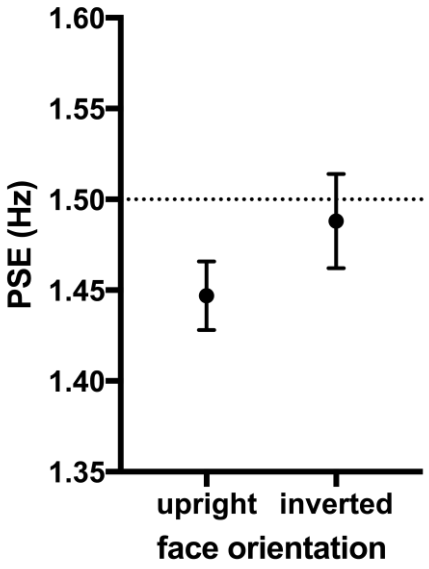


Figure 2.5 Mean PSEs ±1 SEM for the upright and inverted face conditions of experiment 1. Participants judged which of two faces appeared faster, with the standard face oscillating at 1.5Hz and comparison face rate ranging from 0.75 – 2.25Hz.

2.3.4. Discussion

Experiment 1 shows that for upright faces overall facial motion appears slower after adaptation to whole face motion. That it occurred only for upright faces argues against this effect arising merely from low-level visual mechanisms. The finding of upright-specific adaptation effect is unusual – typically, both upright and inverted face adapters produce aftereffects, albeit dissociable by their simultaneous adaptability to opposing stimuli (Rhodes et al., 2004; Webster & MacLeod, 2011). It has been suggested that whereas adaptation to inverted faces adapts neural populations specialised for features (parts), adaptation to upright faces adapts both these and neurons specialised for facial configuration (wholes; Rhodes et al., 2004). Both populations could reside within FFA, explaining this area's responsiveness to faces of either orientation (Haxby et al., 1999 – although a 'neural inversion effect' is possible with careful stimulus design, Kanwisher, Tong, & Nakayama, 1998).

In contrast to the above findings, our novel inversion effect for adaptation could be a result of the dynamic nature of our stimulus. The STS has been implicated in biological motion processing, including facial motion. The relationship between inversion and STS activity is similarly inconclusive however: although inversion of biological motion does not significantly affect STS activation (J. C. Thompson, 2005), TMS over pSTS reduces sensitivity to upright, but not inverted point-light walkers (Grossman, Battelli, & Pascual-Leone, 2005). Using static face stimuli, Yovel & Kanwisher (2005) found a neural inversion effect in face-selective STS at the group level that (unlike FFA) did not correlate with magnitude of behavioural inversion effect across participants. Overall then, the relation of the inversion effect to both adaptation and neural activity is complex, and likely to depend on details of experimental design. Given that this is (to our knowledge) the first combination of adaptation, inversion, and dynamic face stimuli, our orientation-dependent result is not in contradiction to prior literature.

Having reproduced illusory slowing using global facial motion, our subsequent experiments returned to dynamic cross-feature interactions. In pursuit of this, Experiment 2 replaced measurement of perceptual bias with discrimination thresholds, thereby avoiding several theoretical issues inherent to the former.

2.4. Experiment 2: Mouth movement impairs eyebrow discriminability

2.4.1. Introduction

Experiment 2 sought to extend Experiment 1 and Cook et al.'s perceptual results (illusory slowing) on dynamic feature interactions to performance, by measuring feature movement discrimination. Having demonstrated the suitability of our avatar's animation for these purposes, we asked whether the interference of one feature's movement on judgements about another feature (mouth and eyebrows respectively) could manifest in reduced sensitivity as well as a shift in appearance.

The interpretation of PSE shifts entails theoretical ambiguity not suffered by measures of sensitivity. In particular, it is typically not possible to guarantee that a shift from veridical arises from perceptual changes rather than changes in the observer's decision criterion. Though the former is often assumed, it has been shown that experimenter-induced response biases are sufficient to produce equivalent shifts without loss of sensitivity (Morgan, Dillenburger, Raphael, & Solomon, 2012). While there is no specific reason to believe such a confound exists within our Experiment 1 or those of Cook et al., our argument that these cross-feature interactions have their locus in sensory processing would be stronger if we could demonstrate that movement in one feature interferes with sensitivity to movement in another.

We therefore performed a second 2AFC experiment; this time with the task being to detect which face contained a characteristic intra-feature motion signal (misaligned eyebrow movement). Like Cook et al., we manipulated the global motion context with the presence (resp. absence) of concomitant mouth movement, and again we assessed whether any effect could be attributed to face-specific mechanisms by repeating the experiment using upright and inverted stimuli. We hypothesised that sensitivity to the target's misaligned eyebrow movement would decrease in the presence of mouth movement, and in addition that these inter-feature interactions would be apparent for upright, but not inverted faces.

2.4.2. Methods

Participants

There were 10 participants in the study, 6 of whom (including both experimenters) were experienced psychophysical observers. All barring the experimenters were naïve as to the purpose of the experiment.

Stimuli and Apparatus

In each trial, both faces were presented simultaneously. The faces spanned 5 DVA and were situated 1.25 DVA apart. Their size and eccentricity was reduced following Experiment 1 so as to bring their features closer to fixation. Stimuli were again presented against a black background. Following participant response, feedback was provided via text on screen (correct/incorrect), but no trial counter was displayed. Stimuli were presented on a 22in monitor screen 1280 pixels wide by 1024 pixels high, with a refresh rate of 85Hz and at a viewing distance of 54cm. No eye tracker was used.

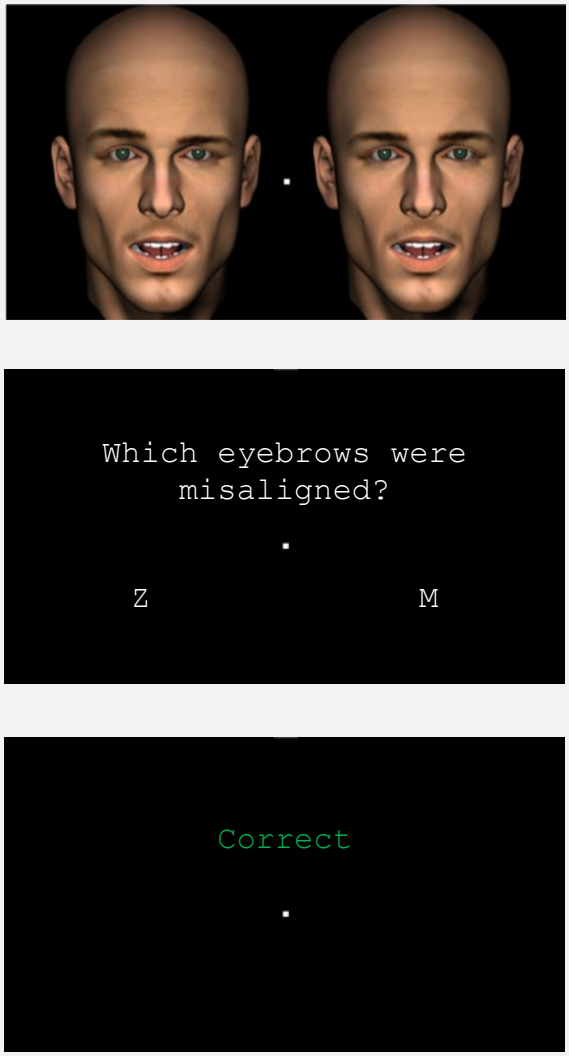
To insert a detectable motion signal into the target face, the broader motion profile of a single feature (its eyebrows) was altered. Specifically, because our facial movement was always periodic and the eyebrows are themselves a feature composed of two parts, a phase-offset in their vertical oscillation could be introduced. This delay between individual eyebrow turning points produces an overall asymmetry (or 'wiggle') that becomes increasingly obvious with increasing phase-offset. We therefore had an intra-feature motion signal whose strength could be varied, permitting sampling across levels for estimation of psychometric functions and detection thresholds.

Design and Procedure

Participants completed a total of 1120 trials split over 4 blocks (with brief rests in between), and the experiment took just under two hours per participant. The four blocks comprised the upright/inverted face and moving/static mouths conditions, separately and fully crossed. In each trial, participants viewed both test faces either side of fixation, both moving at 1.5Hz. In the mouth moving conditions, both faces' mouths oscillated at 0 degrees of phase relative to their eyebrows, such that the mouth opened in time with the eyebrows raising. In the mouth static condition, both mouths remained closed.

The eyebrows of the distractor face were always aligned (i.e. oscillating at 0 degrees of phase relative to each other), while the target face's eyebrow oscillations were misaligned by 1, 5, 10, 15, 20, 25 or 30 degrees of relative phase in randomised trials. The phase of each target face eyebrow 'straddled' its mouth phase, for example with the 5° misalignment level one eyebrow initiated movement at 2.5° while the other started at 357.5°. Which eyebrow led was counterbalanced. 40 measurements per level per orientation per mouth condition were taken (40 × 7 × 2 × 2 = 1120). Both faces' starting phases were 0° – because their speeds were equal and irrelevant to the task, there was no need to offset them relative to each other.

Participants viewed both test faces simultaneously either side of fixation for 3 seconds, before being prompted via on-screen text to indicate, "Which eyebrows were misaligned?" This ISI lasted a minimum of 2 seconds (though again participants typically responded much faster), and for the remainder of the interval red or green text provided feedback as to whether their response was "correct" or "incorrect". This procedure was repeated for all 280 trials of the block.



Test stimuli
3sec
1.5Hz oscillation rate (both faces)
0° relative feature phase (both faces)
or mouths static & closed (both faces)
1°, 5°, 10°, 15°, 20°, 25°, or 30° relative eyebrow phase (target face)
Upright or inverted (both faces)

Inter stimulus interval
2sec +

Figure 2.6. Experiment 2 trial design. Participants were presented with two test faces both of whose mouths oscillated in synchrony with the eyebrows or were stationary and closed. The target face's eyebrows were misaligned by a variable amount. Participants received feedback after providing a response, but no notification of progress.

2.4.3. Results

For each participant, psychometric functions were again fitted as per Figure 2.4 for each of the four combinations of orientation and mouth movement condition. With 50% denoting chance performance, thresholds were then taken as the amount of eyebrow misalignment at which the target was correctly identified 75% of the time.

For upright faces, the mean detection threshold for misaligned eyebrow movement was higher in the presence of mouth movement, and a paired-samples t-test revealed this to be marginally significant (mean difference = 3.11°, *SD* = 4.33°; $t(9) = 2.27$, two-tailed $p = .05$).

For inverted faces however, the mean detection threshold did not differ significantly between mouth moving versus stationary conditions (mean difference = -0.23°, *SD* = 1.92°; $t(9) = -0.38$, two-tailed $p = .71$; see Figure 2.7).

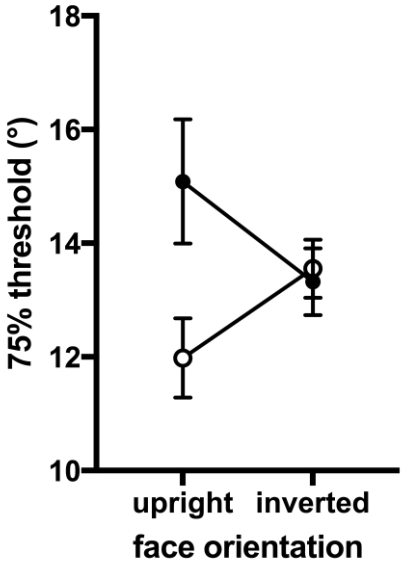


Figure 2.7. Mean thresholds ±1 SEM for the upright and inverted face conditions, when the faces' mouths were moving (solid circles) versus stationary and closed (empty circles). Participants decided which of two faces' eyebrow movement was misaligned, with the target face's eyebrow misalignment ranging from 1 – 30° of relative phase.

2.4.4. Discussion

Although the task required judgement of a single dynamic feature (eyebrows), movement in a task-irrelevant and spatially separate feature (mouths) appears to interfere. Furthermore, the specificity of this interference to upright faces argues against an origin in low-level visual properties such as generic motion (which is preserved across orientations), in favour of face-specific processing. These obligatory cross-feature interactions provide further evidence for a global, holistic encoding of facial movement.

The effect of concurrent mouth movement on sensitivity to eyebrow movement was only marginally significant for upright faces a likely result of higher inter-subject variability (threshold 95% CIs 2.14° and 1.37° for upright mouth moving and static respectively, versus 1.16° and 1.00° for inverted). It is possible that varying difficulty across conditions and stimulus levels could have been a source of confound, if this had differential effects on fixation discipline for example. It is plausible that increased difficulty of some trials could have provoked more lapses in fixation, obscuring our experimental effect and increasing overall noise. Given the addition of mouth movement is hypothesised to increase difficulty for upright but not inverted faces, it is further plausible that such confounds might manifest as increased variability for the upright conditions in particular.

We therefore sought to replicate this experiment while making some adjustments to the design, which were intended to reduce variability.

2.5. Experiment 3: De-correlating eyebrow and mouth movement abolishes mouth movement interference on eyebrow discriminability

2.5.1. Introduction

In an effort to discount the above potential confounds and increase effect size, we repeated the experiment with the addition of an eye tracker to enforce fixation. We also interleaved our conditions more thoroughly in a bid to even out variance. Our hypotheses were again that sensitivity to the target's misaligned eyebrow movement would decrease in the presence of mouth movement, and secondly that these inter-feature interactions would manifest for upright, but not inverted faces.

2.5.2. Methods

Participants

10 participants took part in the study, 7 of whom (including both experimenters) were experienced psychophysical observers. All except the experimenters were naïve as to the purpose of the experiment.

Stimuli and Apparatus

Stimuli were identical to Experiment 2, and were again presented against a black background. Feedback was provided following participant response, and a trial counter indicating progress through the current block was displayed. The monitor screen resolution was 1152 pixels wide by 864 pixels high and participants viewed the stimuli from 57cm away, otherwise the display was as described for Experiment 2. The eye tracker served only to enforce fixation (we were not recording gaze direction or eye movements for analysis); for details see general methods section above.

Design and Procedure

Participants completed a total of 1120 trials split over 20 blocks (with brief rests in between), and the experiment took around 2 hours 45 minutes overall. The increased frequency of breaks was necessary due to the fatiguing effect of using the eye tracker, and its need for regular recalibration. Each block therefore contained 56 trials in which all conditions and levels were randomised except orientation. This meant that along with the amount of eyebrow misalignment in the target face, whether the mouths were moving or stationary varied from trial to trial. Whether faces were presented upright or inverted was randomised across blocks. The same levels of target face eyebrow misalignment were presented as per Experiment 2; we therefore again took 40 measurements per level per orientation per mouth condition ($40 \times 7 \times 2 \times 2 = 1120$).

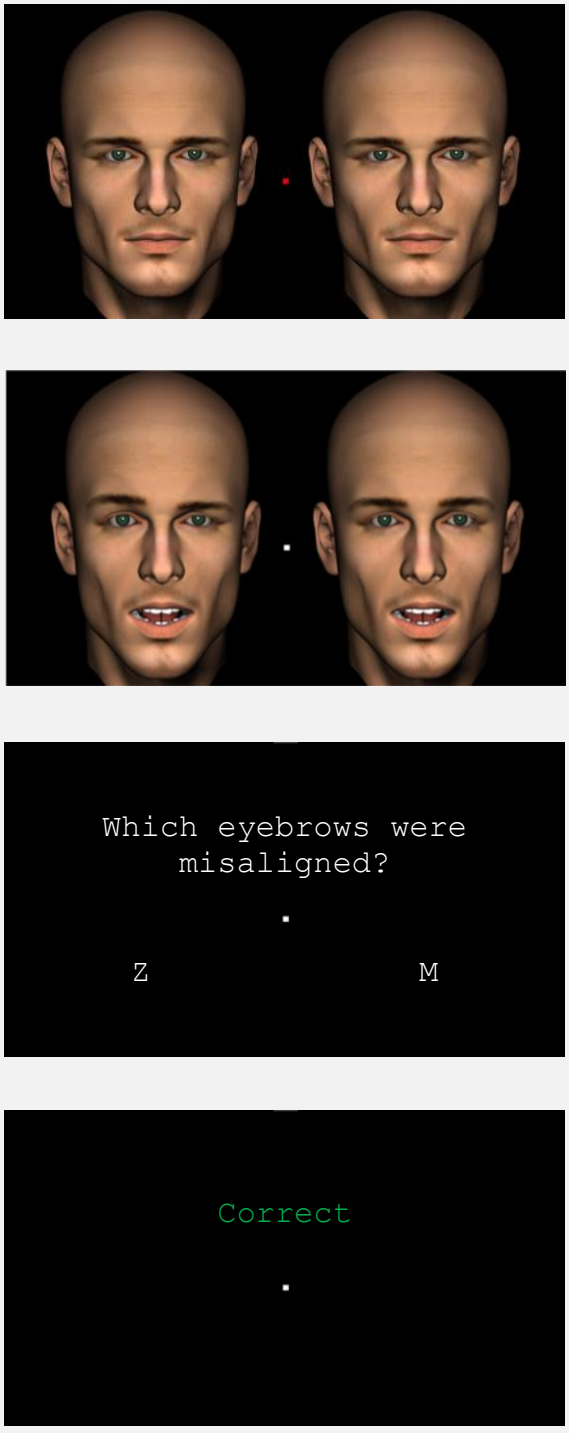


Figure 2.8. Experiment 3 trial design. Trials were as per experiment 2 except that the addition of eye tracking necessitated a drift-correction stage. During drift correction, fixation cycled through red -> green -> white in 200ms steps (red shown).

2.5.3. Results

All trials in which fixation lapses were detected were discarded without requesting a response, and blocks were extended until the required number of responses had been successfully collected. The fitting of psychometric functions for estimation of detection thresholds was as described for Experiment 2. Again we were interested in comparing mean detection thresholds for misaligned eyebrow movement in the presence versus absence of mouth movement, for upright and inverted faces.

For upright faces, there was no difference in mean detection threshold for misaligned eyebrow movement between mouth moving and mouth stationary conditions (mean difference = -0.003°, *SD* = 2.58°; pairwise $t(9) = -0.004$, two-tailed $p = .997$).

For inverted faces, the mean detection threshold also did not differ significantly between mouth moving versus stationary conditions (mean difference = 1.17°, *SD* = 2.56°; pairwise $t(9) = 1.44$, two-tailed $p = .18$; see Figure 2.9).

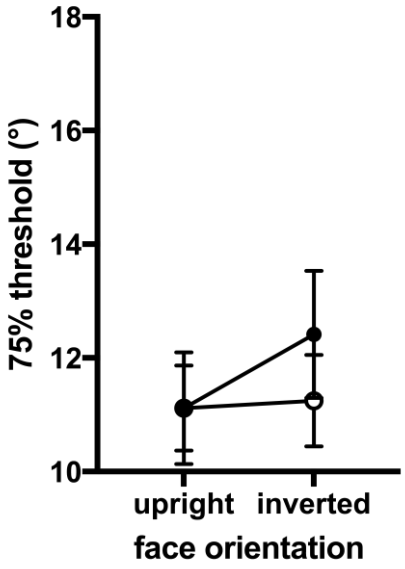


Figure 2.9. Mean thresholds ±1 SEM for the upright and inverted face conditions, when the faces' mouths were moving (solid circles) versus stationary and closed (empty circles). Participants decided which of two faces' eyebrow movement was misaligned, with the target face's eyebrow misalignment ranging from 1 – 30° of relative phase.

2.5.4. Discussion

Experiment 3 failed to find an effect of mouth movement on eyebrow movement sensitivity for either orientation. Furthermore, overall inter-subject variability was comparable to that of Experiment 2 (threshold 95% CIs 1.93° and 1.47° for upright mouth moving and static respectively, versus 2.19° and 1.57° for inverted). Thus although we could guarantee appropriate fixation for every trial, Experiment 2's cross-feature interaction failed to replicate.

Aside from the addition of eye tracking, Experiment 3 also randomised whether the mouths were moving or stationary on a trial-to-trial basis. To the extent that we are seeking to measure joint coding of features arising from their correlated movement however, this could have introduced confound by de-correlating the eyebrows and mouth. In other words, though we sought to measure interference of mouth movement on the eyebrows, randomising the co-occurrence of their movement could have 'trained' participants that the two are in fact dissociable and therefore enabled their judgement based on information from the diagnostic feature alone.

Analogies to this may be found in the static face literature. The whole-face advantage for feature recognition in the part-whole task (see Introduction) is believed to reflect the benefits of holistic processing, but could instead be a product of general contextual 'encoding specificity' (Tulving & Thomson, 1973). That is, recognition of the target feature could be enhanced when presented in the same context as at training (versus in isolation) regardless of whether said context is a face. Tanaka & Farah (1993, Experiment 1) discounted this possibility however by training and testing participants on the target feature within scrambled face contexts. Disrupting holistic cues in this way reversed the effect: participants were better able to recognise features in isolation as opposed to within the scrambled context in which they had been learned. Thus the whole-face advantage in the part-whole task is not a product of generic contextual similarity between training and test, but depends on said context being a coherent face (Tanaka & Simonyi, 2016).

With respect to our Experiment 3, Tanaka & Farah's result shows that it is possible to perceptually dissociate the features of face stimuli by disrupting their spatial coherence, resulting in the loss of subsequently measurable effects resulting from holistic binding. By randomising the co-occurrence (or not) of eyebrow and mouth movement in our design, we argue that temporal coherence has been similarly disrupted, encouraging parts-based judgements and abolishing the dynamic holistic binding that we intended to measure.

Experiments 2 and 3 differed from Experiment 1 in the relative timing of eyebrows and mouth within the test faces. Whereas Experiment 1 set this at 270° – the relative phase at which Cook et al.'s (2015) illusory slowing effect was greatest – Experiments 2 and 3 kept all features bar the individual eyebrows of the target face in synchrony for simplicity. In Experiment 4 we continued to finesse our feature motion discriminability effect by testing across the range of possible feature timings. Importantly, the static mouth control condition therefore comprised a much smaller proportion of trials, and it will be seen (for the 0° condition) that this does indeed result in replication of Experiment 2.

CHAPTER 3. Dynamic feature binding is timing-dependent, and this dependency is lateralised

3.1. Introduction

The first three experiments provided tentative evidence for interdependencies in the perception of features' movement, but we had mixed success in enhancing and delineating this effect. Experiments 4 and 5 continued our attempts to do so by extending our design into feature relative timing and lateralisation.

Experiment 4 replicated the design of Experiments 2 and 3 across the range of possible eyebrow-mouth relative timings, following recent reports that dynamic interactions between eyes and mouth are sensitive to temporal structure. Specifically, this experiment is able to reconcile two superficially contradictory results, demonstrating that comparison against inverted faces and static mouths each provide differing partial views on the overall phenomena. The results also reveal the importance of synchrony across features for their perceptual binding.

Experiment 5 localises this synchronicity-mediated binding to the left cerebral hemisphere, contrasting with what might broadly be expected for face perception but concordant with lateralisation of facial speech processing specifically. The potential inadequacy of 'grand models' of lateralisation for explaining the results is considered, and Experiment 5's failure to demonstrate an inversion effect is discussed in terms of the shift to an interval forced choice design.

3.2. Experiment 4: Eyebrow-mouth perceptual interactions are dependent on relative timing

3.2.1. Introduction

Cook et al. (2015) reported the greatest illusory slowing of eyelid closure when mouth movement either matched or led the eyelid movement (180° and 270°), and these feature relative timings were the only ones in which an inversion effect (greater slowing for upright versus inverted faces) was observed. More recently, Iwasaki & Noguchi (2016) reported a reduction in sensitivity to eye-region micro-expressions in upright faces when concurrent with, or followed by mouth movement, relative to no mouth movement.

If both the illusory slowing and reduction in micro-expression sensitivity are taken as evidence for interference of the mouth on judgements of the eyes (and thus of holistic binding), these results may appear to contradict. Whereas Cook et al.'s effect occurs only when the mouth leads the eyes, Iwasaki & Noguchi's occurs only when the mouth follows the eyes.

There are several methodological differences which may account for this, such as the latter's presentation of Ekman's (1992) basic six expressions versus the former's use of periodic, arguably speech-like movement. The two are somewhat incommensurable in what they measure however, and reconciling them ultimately requires a bridging design that addresses both.

Cook et al. presented differences in the amount of illusory slowing between upright versus inverted faces as evidence for perceptual attraction toward internal models of facial change that have preferred temporal relationships. Importantly then, this account argues against these internal models preferring the timings at which the authors found their effect, with the attraction rather working to increase the perceived delay between opening of the mouth and eyes. It is because their inversion effect is greatest at 180° and 270° that we can discount these timings as being what the putative facial models prefer – otherwise, the stimuli would already match the models' predictions and no attraction would be expected.

Iwasaki & Noguchi presented reduced sensitivity to micro-expressions in the eyes when followed by a smile versus no mouth movement as evidence for backward masking of the eyes by the mouth. Though they used an amodal completion paradigm (Nakayama, Shimojo, & Silverman, 1989) to manipulate holistic binding and confirm the attribution of their overall effect to face-specific processing, they did not compare the holistic versus non-holistic case at each relative timing in the way that Cook et al. did. Similarly, although comparable effects were induced using mouth movements from all six basic expressions in a separate experiment, only that of happiness was tested across the relative timings, despite the target micro-expressions in the eyes deriving also from anger and disgust (i.e. incongruent expressions with respect to the mouth).

In sum, Cook et al. used speech-like periodic movement to find an inversion effect for illusory slowing, which they attributed to perceptual attraction toward internal models. Iwasaki & Noguchi used single-feature components of basic expressions to find a sensitivity reduction, which they attributed to backward masking. That these effects were reported at contrasting eye-mouth relative timings is not contradictory given the above methodological and theoretical differences. Reconciling them has potential to greatly inform our understanding of the role of temporal structure in dynamic face perception, however.

Our Experiment 4 extended the design of Experiments 2 and 3 across the range of eyebrow-mouth relative timings. Inversion was again employed to manipulate holistic binding for all relative phases, and included static (closed) mouth controls for both orientations. We retained the use of an eye tracker to enforce fixation, and full within-trial randomisation of the mouth conditions (i.e. phase relative to eyebrows plus static control). This was despite the potential confound such intermixing introduced to Experiment 3, because with eight mouth moving conditions rather than one, the static mouths trials now constituted only 12.5% of each block rather than 50%.

Like Iwasaki & Noguchi therefore, we were again measuring sensitivity to feature motion, but using periodic speech-like movement similar to Cook et al. The comparison between face orientations at all relative timings affords direct comparability with the latter's results, while the inclusion of static mouths conditions allows comparison with the former's. One might expect that when considering an inversion effect differences in performance at the same phases as Cook et al. would be found (180° – 315°; mouth matching or leading eyebrows), but when comparing moving to stationary mouths differences at similar timing relationships to Iwasaki & Noguchi would be found (mouth concurrent with or trailing eyebrows).

3.2.2. Methods

Participants

There were 16 participants in the study, 9 of whom (including both authors) were experienced psychophysical observers. All barring the experimenters were naïve as to the purpose of the experiment.

Stimuli and Apparatus

Stimuli were identical to those used in Experiment 3, and were again presented against a black background. Feedback and progress through the block following after participant responses. The monitor refresh rate was 60Hz; otherwise the display and viewing conditions were as described for Experiment 3. The eye tracker was utilised as a fixation monitor.

Design and Procedure

Participants completed a total of 720 trials split over 8 blocks (with brief rests in between), and the experiment took around 2 hours per participant. Each block therefore contained 90 trials in which all conditions were randomised except orientation (which was randomised between blocks). Rather than varying the amount of eyebrow misalignment in the target face so as to fit psychometric functions and estimate thresholds, we fixed the misalignment at 12.5° and took percentage correct at this single signal strength as our dependent variable.

Except in the static mouths trials, both test faces' eyebrows and mouths could oscillate at one of eight phase relationships spanning the full possible range in 45° increments. As before and following Cook et al.'s convention, 0° denotes synchronised features where the mouth opens as the eyebrows raise. 45°, 90° and 135° denotes the eyebrows raising before the mouth opens. 180° denotes synchronised features where the mouth opens as the eyebrows lower. 225°, 270° and 315° denotes mouth opening before the eyebrows raise. We presented 40 trials per eyebrow-mouth condition per orientation (40 × [8+1] × 2 = 720).

Within each trial, participants again viewed both test faces simultaneously either side of fixation for 3 seconds, before providing their response in an ISI whose duration was reduced to a minimum of 1 second. The exception was again those trials where the eye tracker detected any lapse in fixation, at which case no response was required and the trial was instead appended to the end of the block for repetition.

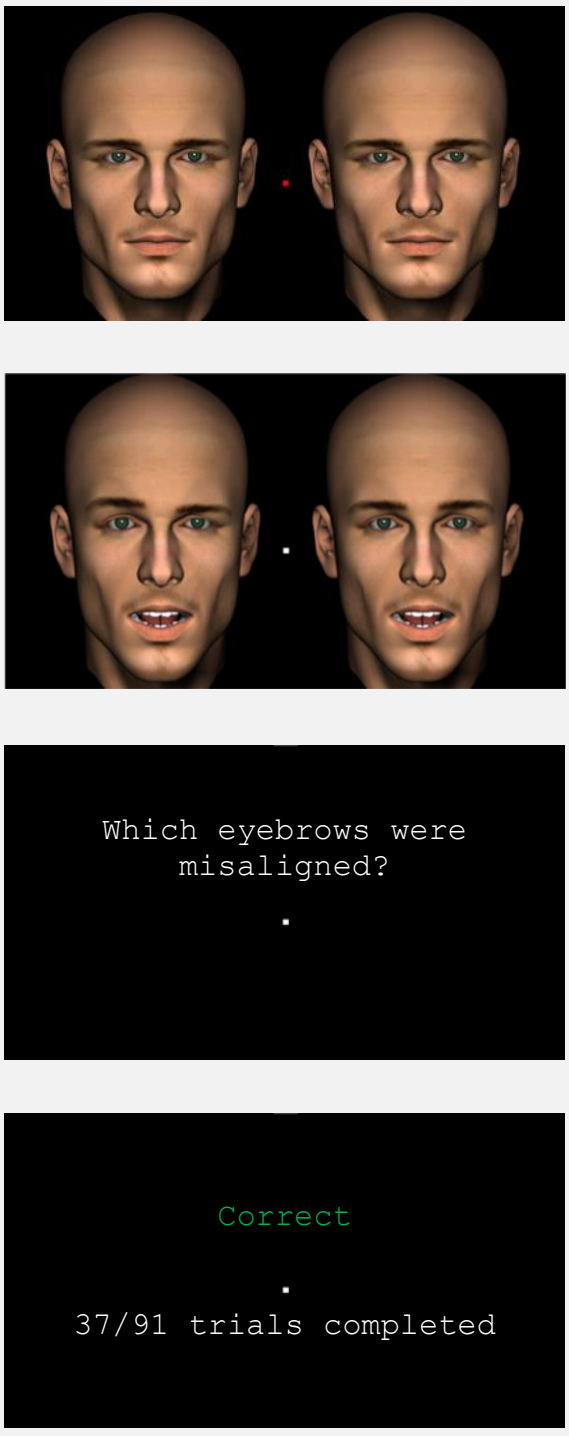


Figure 3.10. Experiment 4 trial design.

3.2.3. Results

For each participant and orientation proportions of correct responses at each eyebrow-mouth relative phase (plus the two static-mouths conditions) were obtained, and means computed (Figure 3.1). Visual inspection indicates firstly that there were differences between mouths following (0° – 180°) as against mouths leading (180° – 360°), and secondly between synchronous (0° and 180°) as against asynchronous (90° and 270°) relative feature motion.

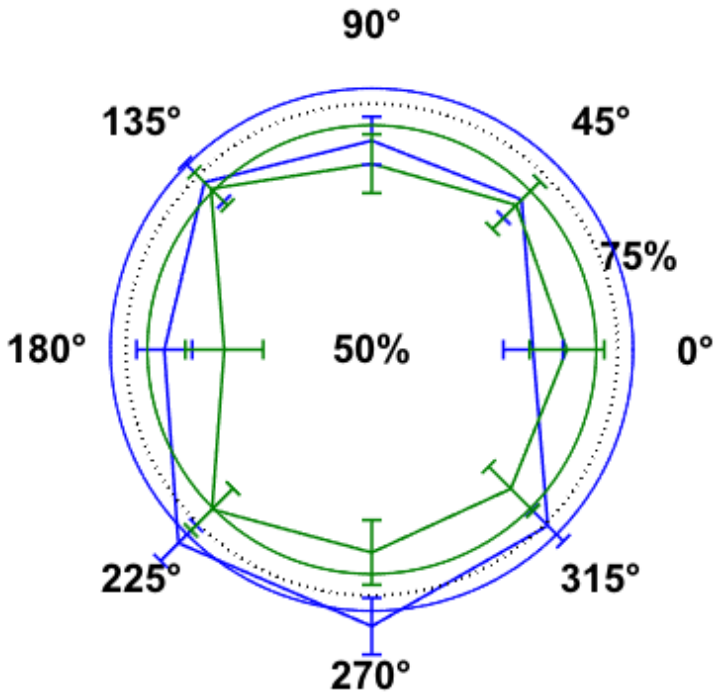


Figure 3.1. Mean percent correct ±1 SEM for the upright (blue) and inverted (green) conditions, when the faces' mouths were moving or stationary and closed. Participants decided which of two faces' eyebrow movement was misaligned, with both faces' eyebrows and mouths oscillating at one of eight phase relationships (plus the static mouth control condition, outer coloured circles). The origin of the polar plot denotes chance performance. the dotted circle denotes 75% correct performance.

Comparing orientations for moving mouths, upright performance can be seen to exceed inverted between 180° – 315° but not between 45° – 135° (with possible reversal at 0°). Conversely when comparing mouth movement conditions for upright faces, the presence of movement (versus static) reduced performance between 0° – 180° but not between 225° – 315°. Overall, the roughly bimodal shape of the moving mouths performance distribution suggests that mouth movement interfered the most at 0° and 180° of feature relative phase. The following analyses serve to confirm the above in relation to the findings of Cook et al., Iwasaki & Noguchi, and our own Experiments 2 and 3.

A 2 × 8 ANOVA on the moving mouths conditions revealed a highly significant main effect of phase ($F(7, 105) = 4.28$, two-tailed $p < .001$), and a significant orientation × phase interaction ($F(7, 105) = 2.34$, $p = .029$; see Figure 3.2). To further explore this interaction, we divided the phases into those in which the mouths matched or led the eyebrows (180° – 315°) versus not, and subjected these collapsed means to a 2 × 2 ANOVA. This revealed a highly significant interaction between orientation and relative timing only ($F(1, 15) = 10.53$, $p = .005$), reflecting that upright performance was significantly better than inverted only when mouth movement matched or led the eyebrows (mean difference for 180° – 315° = 5.94%, $SD = 9.07\%$; pairwise $t(15) = 2.62$, $p = .019$; versus mean difference for 0° – 135° = 0.16%, $SD = 9.46\%$; $t(15) = 0.07$, $p = .948$; see Figure 3.3). These results therefore concord with Cook et al. in relation to which phases support dynamic feature interactions.

We next compared the three possible timing relationships (synchronous; mouths trailing; mouths leading) against the static mouth conditions (see Figure 3.4) for upright faces by collapsing means (0° and 180°; 45° – 135°; 225° – 270°). A one-way ANOVA on the four levels was highly significant ($F(3, 45) = 12.76$, $p < .001$), reflecting that while the effect of mouth movement was highly significant when synchronous with the eyebrows (mean difference = -7.81%, $SD = 5.62\%$; pairwise $t(15) = -5.56$, $p < .001$) and marginally significant when trailing them (mean difference = -4.27%, $SD = 7.92\%$; pairwise $t(15) = -2.16$, $p = .048$). There was no significant effect for mouth movement that led the eyebrows (mean difference = 0.52%, $SD = 7.18\%$; pairwise $t(15) = 0.29$, $p = .776$; see Figure 3.5). These results therefore concord with those of Iwasaki & Noguchi.

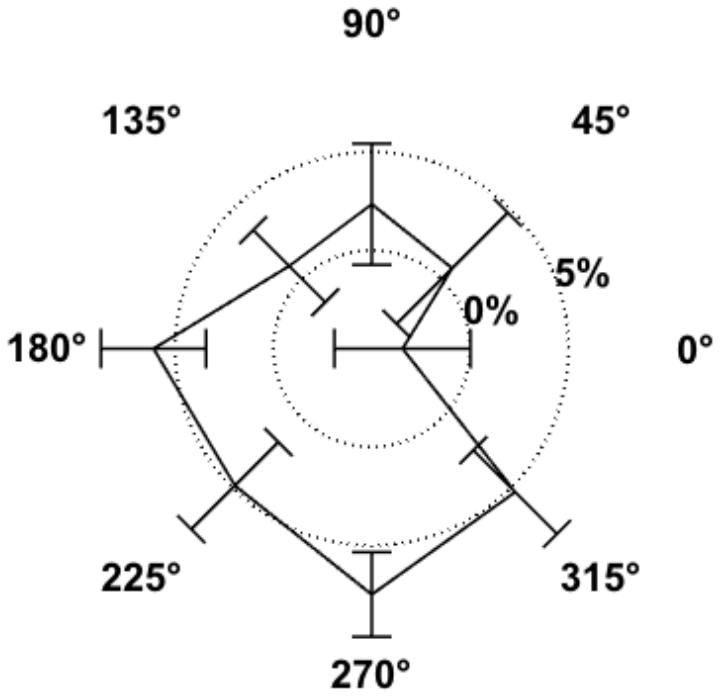


Figure 3.2. Mean difference between upright and inverted face performance ±1 SEM for the moving mouths conditions. The origin of the polar plot denotes a -5% performance difference, i.e. percent correct (detecting which face had misaligned eyebrows) being 5% lower for upright faces.

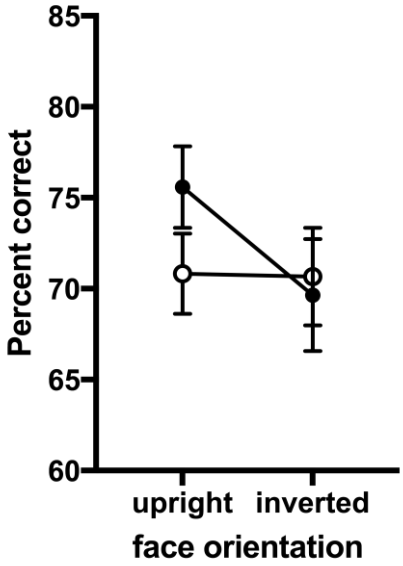


Figure 3.3. Mean percent correct ±1 SEM for the upright and inverted conditions, collapsed over those relative phases where the mouth matched or led the eyebrows (180° - 315°; solid circles) versus not (0° - 135°; empty circles).

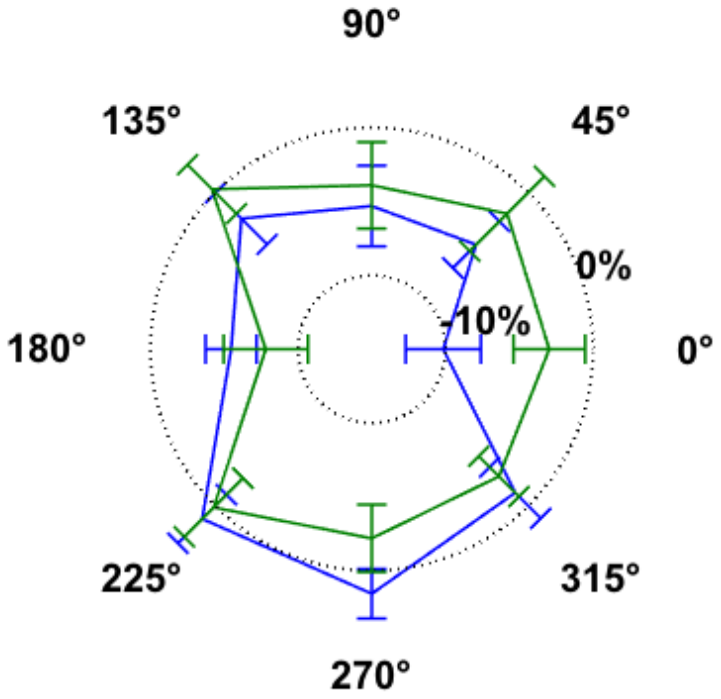


Figure 3.4. Mean difference between mouth moving and mouth static/closed performance ±1 SEM for upright (blue) and inverted (green) faces. The origin of the polar plot denotes a -15% performance difference, i.e. percent correct (detecting which face had misaligned eyebrows) being 15% lower when the mouth was moving at the given relative phase.

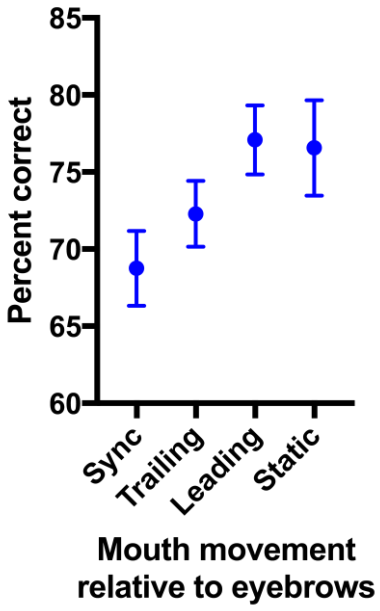


Figure 3.5. Mean percent correct ±1 SEM for upright faces, collapsed over those relative phases where the mouth was synchronous with (0° and 180°), trailing (45° - 135°), or leading (225° - 315°) the eyebrows, and the static/closed mouth condition.

Because the 0° eyebrow-mouth relative phase replicates the design of our Experiments 2 and 3, we repeated their analyses on this single level. For upright faces there was a highly significant reduction in performance when the mouths were moving versus stationary, as revealed by paired-samples t-test (mean difference = -10.16%, *SD* = 10.18%; pairwise $t(15) = -3.99$, $p = .001$). For inverted faces however there was no significant effect of mouth movement (mean difference = -2.97%, *SD* = 9.71%; pairwise $t(15) = -1.22$, $p = .240$; see Figure 3.6). These results therefore concord with Experiment 2.

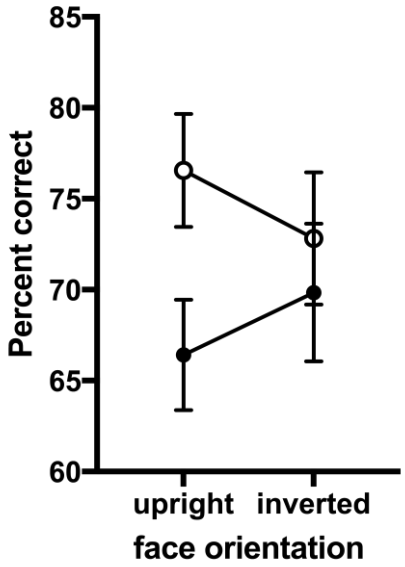


Figure 3.6. Mean percent correct ±1 SEM for the upright and inverted face conditions, when the mouths were moving at 0° of phase relative to the eyebrows (solid circles) versus stationary and closed (empty circles).

When mouth movement matches or leads the eyebrows (180° – 315°) therefore, performance with upright faces is similar to that of the stationary/closed mouths case. When mouth movement follows the eyebrows however (45° – 135°), performance drops from this level toward that of the inverted face case. The reduction of performance is especially severe for upright faces at 0° (mouth movement opposing eyebrow movement).

Finally, to more directly assess the role of synchrony (i.e. the apparent bimodality of the data), we subtracted the upright static mouth performance from each of the upright moving mouths conditions and collapsed over these "effect of mouth movement" difference scores so as to compare synchronous (0° and 180°) against asynchronous (90° and 270°) relative feature timings. The effect of mouth movement on eyebrow movement sensitivity was greater for synchronous relative timings (-7.81%, *SD* = 5.62% versus -1.88%, *SD* = 7.32%), and this synchrony-dependency was highly significant (pairwise $t(15) = -3.57$, $p = .003$; Figure 3.7).

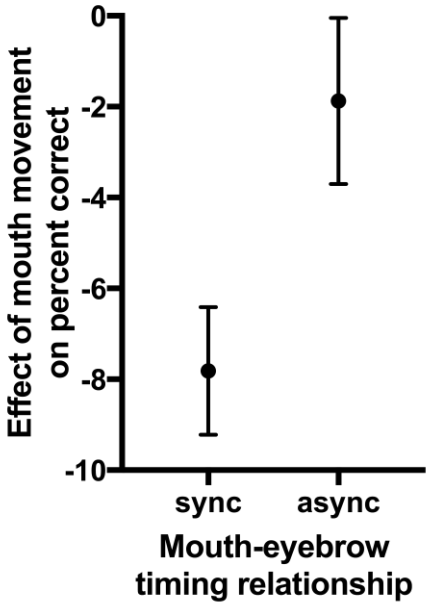


Figure 3.7. Mean difference between mouth moving and mouth static/closed performance ±1 SEM, collapsed over those phases where the features were synchronous (0° and 180°) versus the phases directly orthogonal to synchrony (90° and 270°).

3.2.4. Discussion

Experiment 4 demonstrates that Iwasaki & Noguchi's effect is a product of loss of performance advantages for upright faces (those reported by Cook et al.) when the mouth trails, and an especial reduction in performance when feature movement is synchronous (as found in our Experiment 2). Unlike Experiment 3, Experiment 4's highly significant effect of mouth movement at 0° lends weight to the argument that the former failed to replicate because of its high proportion of randomly interleaved static mouth trials.

Experiment 4 also underscores the complexity of dynamic feature binding's dependence on relative timing, which inversion effects alone cannot fully capture. It does not appear that Cook et al.'s perceptual attraction away from veridical should be equated with 'interference' of the kind that would be expected to reduce performance for example, given the inversion effect at the same relative feature timings consists of a performance advantage for upright faces. With respect to mouth movement induced performance reduction, this was found to be greatest for synchronous feature movement rather than mouth trailing eyes in contrast to Iwasaki & Noguchi.

In presenting our final behavioural experiments employing visual adaptation, Chapter 4 below considers the possibility that perception at asynchronous feature timings may be attracted toward internal models preferring synchrony. Another distinction between our paradigm and that of Iwasaki & Noguchi is the type of facial movement, however. Whereas theirs were formed from static images of basic expressions to which featural change was applied by morphing, ours consisted of smooth and periodic speech-like motion of a CGI avatar. Experiment 5 examined whether the timing dependency of our observed cross-feature motion encoding is lateralised to one cerebral hemisphere. Whereas face perception is widely regarded as right-lateralised, language processing is a canonical example of left-lateralisation.

3.3. Experiment 5: The timing dependency of eyebrow-mouth perceptual interactions is specific to the right visual hemifield

3.3.1. Introduction

A substantial body of experimental and neuroscientific research holds that face perception is lateralised to the right hemisphere (hereafter RH). Divided visual field experiments (hereafter DVF, reviewed by Banich, 2003) form much of the early evidence for this, whereby face stimuli are presented either side of fixation and reaction time or accuracy compared (Leehey, Carey, Diamond, & Cahn, 1978; Rizzolatti, Umilta, & Berlucchi, 1971). Such designs are predicated on the connection of each nasal hemiretina to the contralateral optic tract (and each lateral hemiretina to the ipsilateral tract), so that information from each visual field is received first by the contralateral cerebral hemisphere. Excepting corpus callosotomy (as received by 'split-brain' patients, Gazzaniga, 2000), it is understood that both hemispheres ultimately gain access to both visual fields (hereafter L/RVF), so the DVF paradigm is believed to tap into lower latency and/or more direct connectivity enjoyed by each hemisphere for its respective visual field.

DVF experiments have shown that recognition memory is better in the LVF as might be expected from a signal detection theoretic approach (Finlay & French, 1978), that manual (but not verbal) response times for identifying faces are faster in the LVF (Geffen, Bradshaw, & Wallace, 1971), and that this differential reaction time develops in later childhood (Reynolds & Jeeves, 1978, but see conversely Broman, 1978). The story is not entirely clear-cut however, with some studies failing to find a differential inversion effect between the two visual fields (Bradshaw, Taylor, Patterson, & Nettleton, 1980; Ellis & Shepherd, 1975).

Neuroscientific evidence includes prosopagnosia acquired following right-lateralised brain damage (De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994; Wada & Yamamoto, 2001) and neuroimaging (Kanwisher, McDermott, & Chun, 1997; Sergent, Ohta, & Macdonald, 1992) . From the latter however, results are again somewhat mixed both with regard to inversion, and the extent of FFA's lateralisation specifically. A comprehensive review of neuroimaging studies concluded that FFA is likely to be 'generally' rather than 'absolutely' right-lateralised (Dien, 2009). It could be that face-sensitive regions within the two hemispheres are responsible for different aspects of face representation, and that bilateral interhemispheric modulation at earlier levels of the face-processing hierarchy (left OFA modulating right OFA) drives subsequent and overall right-biased activity (Frässle et al., 2016).

A large body of evidence supports the left-lateralisation of language (for reviews, see Behrmann & Plaut, 2015; Price, 2012). As a result, face and language perception are often juxtaposed when reviewing lateralisation as an overall neuroanatomical organisational principle (Behrmann & Plaut, 2015; Dien, 2009), and interdependencies in their development have long been suspected. It is plausible that specialisation for faces, emerging later in development, is consigned to the RH by the earlier left-lateralisation of language (Behrmann & Plaut, 2015). Regarding visual language (reading) specifically, predisposition of the left hemisphere (hereafter LH) toward higher spatial frequencies (Sergent, 1982) may also be involved, though both the consistency of supporting results (Dien, 2008) and direction of causality (Ossowski & Behrmann, 2015) have been questioned.

Facial speech and lipreading lie at the junction of these abilities, with the fundamentally cross-modal nature of speech comprehension illustrated by the McGurk effect (McGurk & MacDonald, 1976). Within face perception, early case studies indicated dissociability between lipreading and expression or identity recognition (Campbell, Landis, & Regard, 1986), showing that prosopagnosia resulting from unilateral damage does not impair face processing universally.

Neuroimaging studies of speechreading (a more recent term reflecting the involvement of the whole face in conveying meaning) report either bilateral or left-lateralised brain activity. This contrasts with right-lateralisation of the more extensively studied types of face perception described above, while being commensurate with the left lateralisation of language.

Calvert et al. (1997) presented participants in the scanner with silent facial speech movements, non-English but speech-like movements, and non-speech 'gurning' movements. In all but the latter condition increased activity in bilateral auditory cortex was observed, demonstrating that while these areas will respond to non-auditory facial input, it has to be at least 'speech-like' in nature.

Calvert & Campbell (2003) extended these results by presenting static frames of faces enunciating monosyllables at their apex, and when contrasted against resting faces (superposed with the letter "V") activation in left-lateralised Broca's area, dorsal Wernicke's area, and STS was observed, alongside other regions implicated in biological motion. Dynamic speech sequences differentially activated the same regions as well as many more and with greater magnitude, particularly visual motion areas such as V5/MT+.

More recently, De Winter et al. (2015) conducted an extensive fMRI experiment presenting dynamic facial expressions to humans and macaques. Of particular interest is the inclusion of a facial speech condition. In humans, the left pSTS showed relatively higher activation for speech movement over chewing or fear expressions, and whereas activation to these latter two stimuli was consistently right-lateralised, there was no overall lateralisation for speech movements. The authors postulated that whereas the right pSTS is engaged by all dynamic facial representations, the left pSTS is responsible for prelexical visual cue integration resulting in bilateral responses only to facial speech.

Following informal reports by our participants of differential ability to detect the target face's eyebrow misalignment in the two visual fields, we adapted our experiment to a DVF design. By replacing simultaneous presentation of target and distractor with two-interval forced choice in alternating hemifields, we sought to replicate Experiment 4 while assessing the potential lateralisation of our timing-dependent eyebrow-mouth interaction. To the extent that the animation of our stimuli is 'speech-like', one might expect greater modulation by relative timing in the RVF, while the role of inversion in DVF experiments is more ambiguous.

3.3.2. Methods

Participants

There were 22 participants in the study, including both authors who are experienced psychophysical observers. All barring the experimenters were naïve as to the purpose of the experiment. Handedness of participants was not recorded.

Stimuli and Apparatus

Stimuli, feedback and progress counter, display, viewing conditions, and use of eye tracker were described for Experiment 4.

Design and Procedure

Participants completed a total of 640 trials split over 20 blocks (with brief rests in between), and the experiment took around 3 hours per participant overall. Each block therefore contained 32 trials in which phase was randomised and visual field of presentation alternated between left and right. Orientation was again randomised between blocks. As in Experiment 4 we fixed the target eyebrow misalignment at 12.5° and took percentage correct at this single signal strength as our dependent variable.

We did not include a static mouths condition, and both test faces' eyebrows and mouths could oscillate at one of four phase relationships: 0°, 90°, 180°, or 270°. We presented 40 trials per eyebrow-mouth condition per visual field per orientation ($40 \times 4 \times 2 \times 2 = 640$).

Within each trial, participants viewed both test faces in series on one side of fixation for 3 seconds each, before indicating which contained misaligned eyebrows in an ISI lasting a minimum of 1 second. The exception was again those trials where the eye tracker detected a lapse in fixation, in which case no response was requested and the trial was instead appended to the end of the block for repetition. Whether the target or distractor face came first was randomised and balanced across trials, to prevent the use of presentation order as a cue by participants.

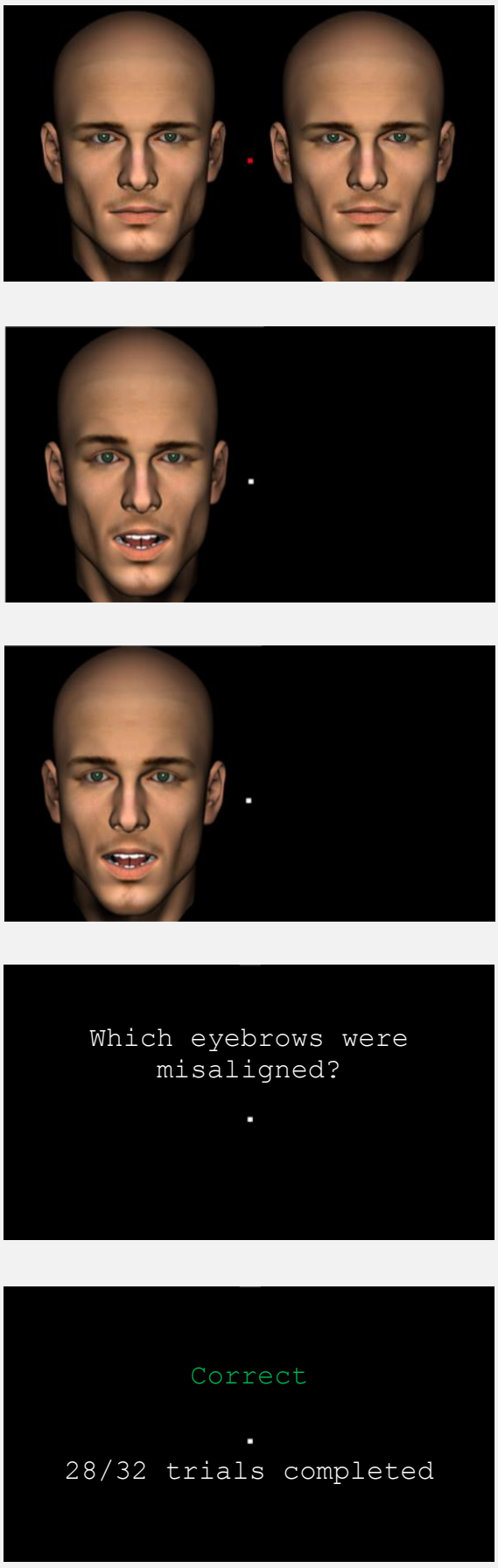


Figure 3.8. Experiment 5 trial design. Test stimuli were presented sequentially in a single visual field (which alternated between trials).

3.3.3. Results

For each participant, orientation, and visual field proportions of correct responses at each eyebrow-mouth relative phase were obtained, and means computed (see Figure 3.11 and Figure 3.10).

A 2 × 2 × 4 ANOVA revealed a significant main effect of phase ($F(3, 63) = 4.05$, two-tailed $p = .011$), and a highly significant visual field × phase interaction ($F(3, 63) = 4.32$, $p = .008$). There was no significant effect of orientation, however ($F(1, 21) = 2.58$, $p = .123$).

To further explore the visual field × phase interaction, we collapsed over orientations and phases so as to compare synchronous (0° and 180°) against asynchronous (90° and 270°) relative feature timings. This revealed a highly significant main effect of synchronicity ($F(1, 21) = 9.41$, two-tailed $p = .006$), and a significant synchronicity × phase interaction ($F(1, 21) = 5.24$, $p = .033$). As Figure 3.9 shows, this interaction reflects a selective reduction in performance for synchronous feature motion in the right visual field.

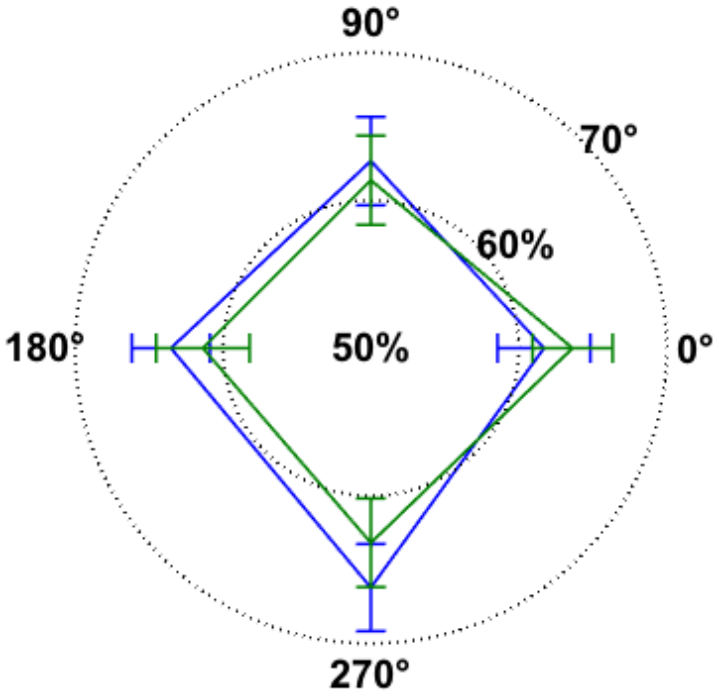


Figure 3.11. Mean percent correct ±1 SEM for the upright (blue) and inverted (green) conditions, when the faces were presented in the left visual field.

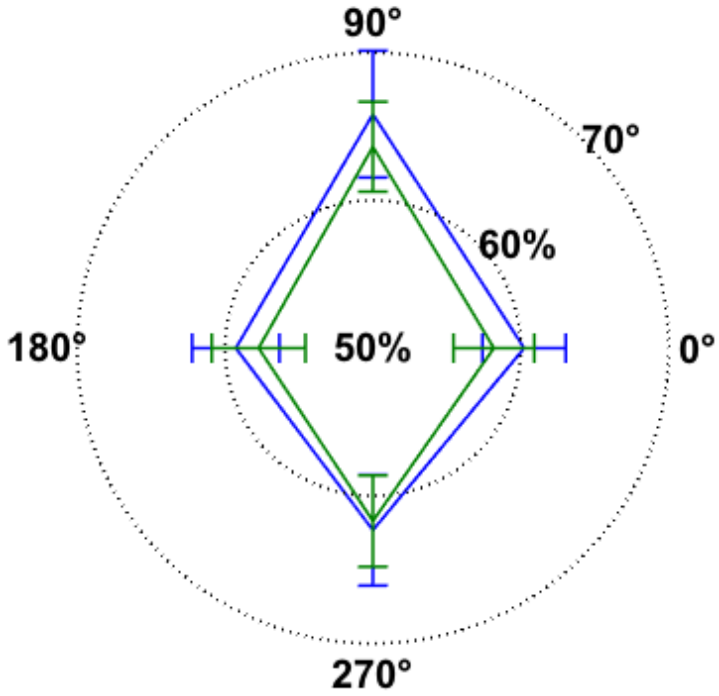


Figure 3.10. Mean percent correct ±1 SEM for the upright (blue) and inverted (green) conditions, when the faces were presented in the right visual field.

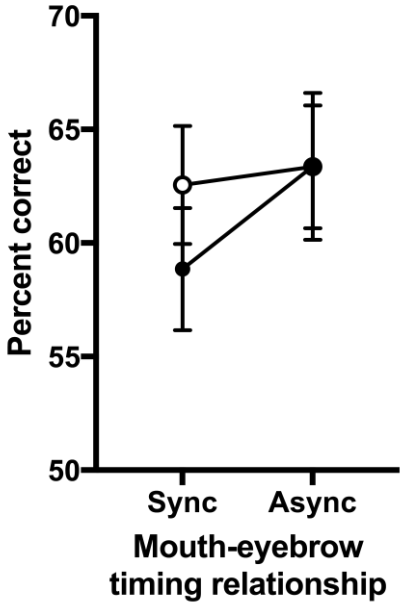


Figure 3.9. Mean percent correct ±1 SEM for the left (empty circles) and right (solid circles) visual fields, collapsed over synchronous (0° and 180°) and asynchronous (90° and 270°) relative feature phases.

3.3.4. Discussion

The central role of synchrony for our cross-feature interference effect has been replicated and localised to the left cerebral hemisphere. This constitutes further evidence for the involvement of the language-specialised hemisphere in processing of facial speech, and contrasts with face perception's traditional right lateralisation.

Specifically, the RH seems equally sensitive to eyebrow movement whatever its timing relative to the mouth. The LH can match the RH's eyebrow sensitivity when mouth and eyebrow turning points do not coincide, however when the features move in synchrony the two are perceptually bound such that sensitivity to the individual feature's motion is impaired.

It is worth considering our cross-feature binding in the context of another proposed hemispheric dichotomy – that of 'analytic' versus 'holistic' cognitive style. This model (Bradshaw & Nettleton, 1981) holds that the LH possesses 'sequential' processing mechanisms emphasising temporal structure and order, whereas the right possesses 'parallel' mechanisms emphasising spatial structure and Gestalt. The heavy influences of language and face perception on this model are clear, but it is less clear how our temporal structure dependent holistic binding effect might fit into it. We appear to have an effect that entails both 'cognitive styles', so a priori the analytic/holistic model would seem to predict lateralisation both ways. It may simply be that such broad models of lateralisation sit at too high an explanatory level to be appropriate for the nuance of dynamic facial speech.

Unlike Experiments 2 and 4, no inversion effect for either hemisphere was found. As has been mentioned the relationship of face inversion to lateralisation is not completely understood, and results could hinge on the subject strategies a particular design allows (Dien, 2009). Our Experiment 5 was an exact replication of Experiment 4 barring the sequential presentation of both test faces in the same region of visual space. If comparison of the faces is necessarily serial, then sequential presentation makes the task much easier; similarly having 3 seconds to view each face as opposed to both effectively doubled the amount of information with which a decision could be made. Sequential presentation also necessitates differential comparison against memory, which may have greater consequences for participant responses than is often assumed (Nachmias, 2006). The phase-dependency of the effect discounts low-level motion or attention-based explanations however, because the results require a specific temporal relationship between the features.

Experiments 6 and 7 continued to probe the neural encoding of our effect. Using adaptation, we sought to clarify the involvement of synchrony, binding, and perceptual attraction.

CHAPTER 4. Adaptation to facial motion reveals the importance of synchrony

4.1. Introduction

Chapter 3 demonstrated that modulation of sensitivity to eyebrow motion by concurrent mouth motion depended on the relative timing of these features. The interference was particularly strong for synchronous feature motion, an effect that was lateralised to the left cerebral hemisphere. The final two experiments explored the special role of synchrony using visual adaptation.

Experiment 6 asked whether different relative feature timings are coded by distinct sensory channels, including whether synchronous temporal relationships shared a common channel. Surprisingly, a general reduction in performance following adaptation to asynchrony only was found. Two potential explanations are discussed.

Experiment 7 tested one of these explanations, that facial timing adaptation should instead be regarded as recalibration of synchrony-centred internal models. By measuring the build up of adaptation over time, we sought evidence for perceptual shifting of near-synchronous facial movement toward synchrony. Although specific analyses revealed some shifting worthy of further study, our alternative account – that feature-wise adaptation is gated by dynamic feature incorporation into a holistic percept – would seem the more fruitful future direction.

4.2. Experiment 6: Adaptation to asynchronous facial motion impairs eyebrow discriminability irrespective of relative timing at test

4.2.1. Introduction

In reconciling two recent findings on eye-mouth perceptual interactions, Experiment 4 also revealed a special role for synchrony in the binding of dynamic facial features. The dependency of this binding on relative feature timing could be a product of different sensory channels coding for the various timing relationships. For example, if synchronous facial movement (0° and 180° of eyebrow-mouth relative phase in our paradigm) is coded by a common channel, then adaptation to 0° might be expected to reduce sensitivity both at this phase, and at 180°.

Sensory adaptation has been studied since the earliest days of psychophysics. Helmholz (1910, as cited in Fernández-Ruiz & Díaz, 1999) deployed prismatic goggles to study visuomotor plasticity, showing that whereas a constant offset in visual position induced by prisms initially produces concomitant errors in motor accuracy, this error reduces over successive attempts. The 'correction' of the nervous system for an introduced perturbation involves realignment of visual and proprioceptive spatial maps, which – due to persistence of the realignment – causes 'rebound' upon removal of the goggles, i.e. gradually reducing motor errors in the opposite direction.

The motor aspect of the prism adaptation effect is important, given the effect's magnitude increases with number of attempted motor acts rather than length of time wearing the goggles (Fernández-Ruiz & Díaz, 1999). Many such adaptation effects have been documented in the purely visual domain however, with the broad principles of recalibration and subsequent rebound continuing to apply. In the classic motion aftereffect or 'waterfall illusion' (Addams, 1834) for example, fixating one's eyes on a coherently moving pattern for several tens of seconds will cause a subsequently viewed static pattern to appear to drift in the opposite direction.

The ubiquity of adaptation throughout sensory systems (Cortes et al., 2012) renders it a powerful tool for the study of their underlying computational mechanisms, often being referred to as the 'psychophysicist's electrode'. Alongside perceptual aftereffects, adaptation often entails reduction of sensitivity to the adapting stimulus, which can be used to make inferences about the neural division of labour for encoding that stimulus class. Orientation specific contrast elevation provides an illustrative example (Blakemore & Campbell, 1969), given the similarly circular nature of our facial timing independent variable. That adaptation to a grating elevates contrast detection thresholds for subsequently presented gratings of similar, but not different orientations is classic evidence of the adapting and test orientations' encoding by distinct neurons (firing rates of the former having been selectively reduced, Schwartz, Hsu, & Dayan, 2007) . The specificity of adaptation-induced sensitivity reduction therefore speaks to the underlying neural representation of stimulus dimensions.

Regarding face perception, channel-based coding corresponds to the less popular of the competing exemplar and norm-based models (see general introduction above). Most adaptation studies have used static stimuli and considered adaptation aftereffects (bias) rather than sensitivity (threshold; Schwartz et al., 2007), with the results typically regarded as favouring the norm-based account (Rhodes & Leopold, 2011). Repulsive aftereffects for dynamic expressions have similarly been explained via norm-based face space (Cristóbal Curio et al., 2008), however several have pointed out that an exemplar-based explanation also accommodates many of these findings (Ross et al., 2014; Storrs & Arnold, 2015). Overall then, while a norm-based account seems more likely the argument has yet to be settled psychophysically, and the applicability of existing research to relative feature timing as emphasised by our periodic stimuli is limited. Given the aforementioned analogies to orientation and motion direction perception, probing for channels for facial timing using adaptation effects on sensitivity remains a reasonable new direction.

To do so, we again tested participants' sensitivity to eyebrow misalignment across the range of eyebrow-mouth relative phases. This time however, testing was preceded by adaptation to those same relative phases in a fully crossed design. If different relative timings are encoded by activity in distinct neural populations (channels), then prior adaptation to one phase should selectively reduce sensitivity at that same phase. Alternatively, if multiple phases share a common channel, then this should reflect in cross-adaptation. In statistical terms, these predictions would manifest as an interaction between adapting and test phases.

4.2.2. Methods

Participants

15 participants took part in the study, 8 of whom (including both authors) were experienced psychophysical observers. All barring the experimenters were naïve as to the purpose of the experiment.

Stimuli and Apparatus

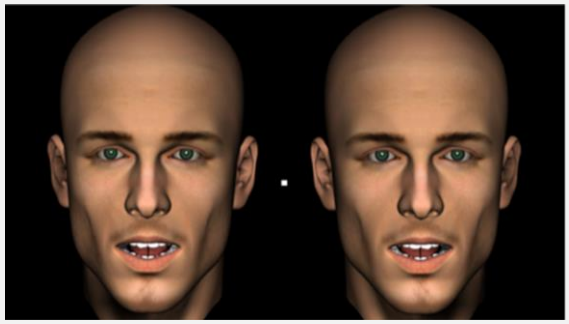
Stimuli, feedback and progress counter, display, viewing conditions, and use of eye tracker were described for Experiment 4.

Design and Procedure

Participants completed a total of 800 trials split over 5 blocks (undertaken at participants' convenience), and the experiment took around 3 hours 15 minutes per participant overall. Each block therefore contained 160 trials in which test phase was randomised. Adapting condition was randomised between blocks, and only upright faces were presented. As in Experiment 4 we fixed the target eyebrow misalignment at 12.5° and took percentage correct at this single signal strength as our dependent variable.

During adaptation both test face positions contained identical adapting faces, whose eyebrows and mouths oscillated at one of four phase relationships: 0°, 90°, 180°, or 270°. Neither adapting face contained misaligned eyebrows. We also included a fifth control condition without adaptation. During presentation of test stimuli, both faces' eyebrows and mouths could similarly oscillate at one of 0°, 90°, 180°, or 270° phases (the target face contained eyebrow misalignment as above). We presented 40 trials per adapting condition per test condition (40 × [4 + 1] × 4 = 800).

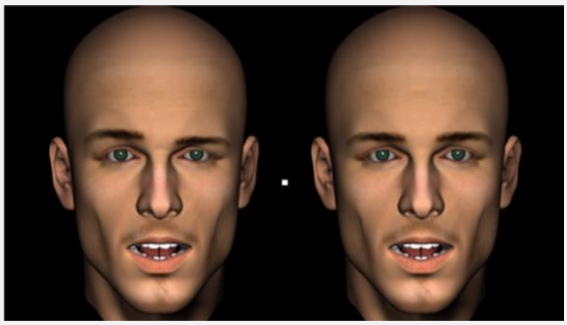
Within each trial, participants viewed both adapting faces simultaneously both sides of fixation for 5 seconds (or 30 seconds at the start of the block), followed by eye-tracker drift correction as in Experiments 3 and 4. During this brief drift-correction stage, a static pair of faces occupied the stimuli positions, and the subsequent onset of test stimuli was signalled by a sequence of fixation colour changes as in earlier experiments. Both test stimuli were then presented simultaneously for 2 seconds, before a final 1 second (minimum) ISI in which participants gave their response and received feedback. The exception was again those trials where the eye tracker detected any lapse in fixation, in which case no response was required and the trial was instead appended to the end of the block for repetition.



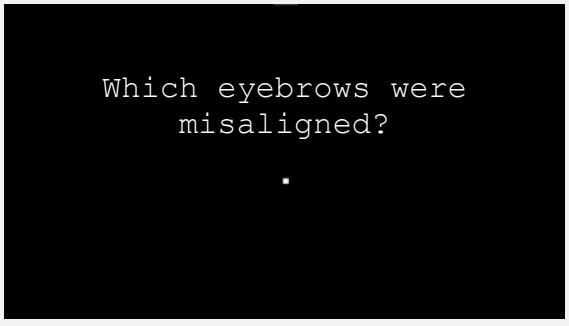
Adaptation (except no adapt condition)
30sec initial / 5sec top-up
1.5Hz oscillation rate (both faces)
0 – 270° relative feature phase (90° steps; both faces)
Upright (both faces)



Eye tracker drift correction
600ms +



Test stimuli
3sec
1.5Hz oscillation rate (both faces)
0 – 270° relative feature phase (90° steps; both faces)
12.5° relative eyebrow phase (target face)
Upright (both faces)



Inter stimulus interval
1sec +

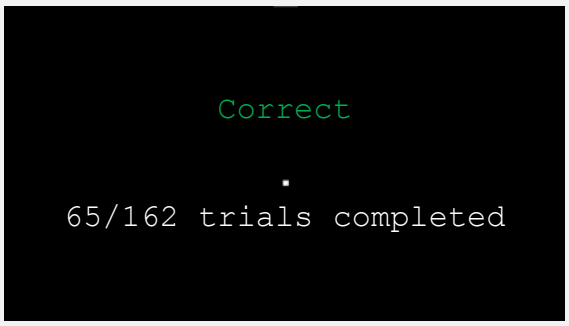


Figure 4.1. Experiment 6 trial design. Adaptation relative feature phase varied between blocks whereas test feature phase varied between trials. The no adaptation condition simply omitted each trial's initial adaptation stage.

4.2.3. Results

For each participant and adapting condition proportions of correct responses at each eyebrow-mouth relative phase were obtained, and means computed (Figure 4.2). Visual inspection indicates subdivision of both the adapting and test phases into synchronous versus asynchronous relative timings. Reduced performance was seen following adaptation only to the 90° or 270° phases (relative to the no adapt control), and generally participants performance was worse when tested at 0° or 180°. The following analyses serve to confirm the above in terms of our results from Experiments 4 and 5.

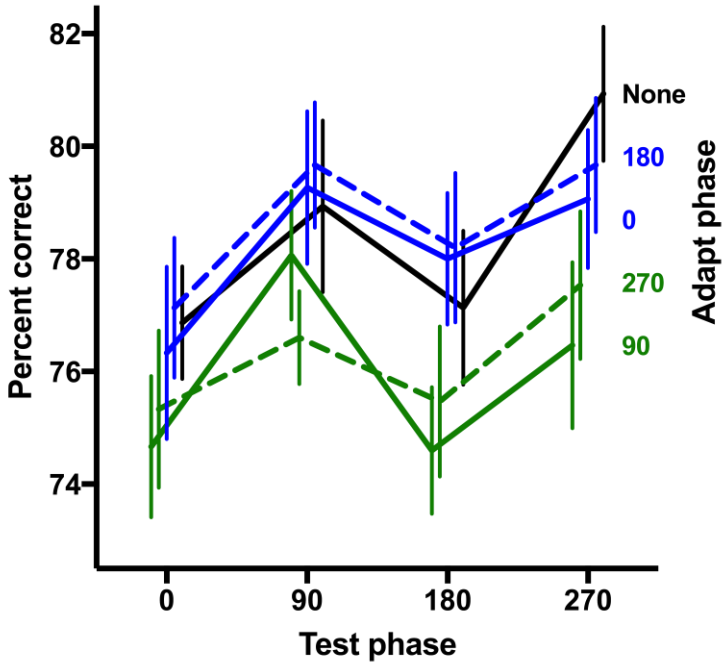


Figure 4.2. Mean percent correct ±1 SEM for the four test face feature relative phases, following adaptation to the same phases (or no adaptation in the control condition).

A 5 × 4 ANOVA revealed a highly significant main effect of adapting condition, $F(4, 56) = 6.90$, $p < .001$ and a highly significant effect of test relative phase, $F(3, 42) = 15.76$, $p < .001$. Their interaction was non-significant however, $F(12, 168) = 0.81$, $p = .645$.

To assess the role of synchrony we collapsed over 0° and 180°, and over 90° and 270°, for both the adapting and test independent variables (Figure 4.3). A 3 × 2 ANOVA (adapt: synchronous, asynchronous, or none; test: synchronous, or asynchronous) revealed a highly significant main effect of adapting synchronicity, $F(2, 28) = 12.37$, $p < .001$, and a highly significant effect of test synchronicity, $F(1, 14) = 28.04$, $p < .001$, but again no interaction, $F(2, 28) = 0.50$, $p = .61$. Post-hoc comparisons on the collapsed adapt synchronicity levels showed that while there was no significant difference between the synchronous adapt and no adapt conditions, there were highly significant differences between both of these and the asynchronous adapt condition (see Table 4.1).

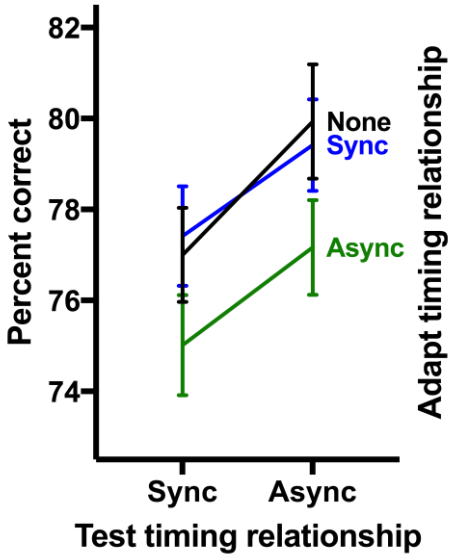


Figure 4.3. Mean percent correct ±1 SEM collapsed over synchronous (0° and 180°) and asynchronous (90° and 270°) relative feature phases for both test and adapting conditions.

	Mean	*SD*	*t*	*df*	*p*
Sync - None	-0.05	2.26	-0.09	14	1
Sync - Async	2.33	1.83	4.93	14	.001
None - Async	2.38	2.23	4.13	14	.003

Table 4.1. Pairwise comparisons of the collapsed adapting conditions, after additionally collapsing over test relative feature phase (so as to compare only the main effect of adaptation). *p* values are Bonferroni corrected by multiplying by 3.

4.2.4. Discussion

Figure 4.2 shows that performance is generally better at 90° and 270° of relative phase, and worse following adaptation to these phases. By contrast, performance following adaptation to 0° and 180° is similar to the no adaptation condition. It appears that while performance for detecting misaligned eyebrow movement is best in the context of asynchronous global motion (i.e. eyebrows leading mouth or vice versa such that turning points do not coincide), it is also most impaired by adaptation to global asynchrony (Figure 4.3). The lack of adapt-test relative phase interaction reflects that asynchronous adaptation impairs performance across the board, arguing against a simple channel-based account.

That the adaptation effect is feature relative phase dependent argues against its origin in lower-level visual mechanisms (e.g. generic motion adaptation), which are not sensitive to such spatially distal temporal relationships. Furthermore, the lack of adaptation specifically to synchronous motion is unusual: participants' sensitivity to eyebrow misalignment appears to be unaffected by it. Across adapting conditions, the pattern of greater interference of synchronous mouth movement on the eyebrows at test replicates our findings from Experiments 4 and 5, supporting the theory that synchronous feature motion promotes binding. The question then is why those relative feature timings that reduce sensitivity to single-feature motion fail to induce further performance reductions by adaptation. Note that this is not a floor effect – performance does drop further following adaptation to facial asynchrony.

Rather than an adapt-test phase interaction (indicative of separate channels), we found a reduced sensitivity to eyebrow motion at all test phases following asynchronous adaptation. This is consistent with local adaptation to the single feature, whose misalignment needs to be similarly detected across all test phase conditions. That sensitivity to eyebrow misalignment was unimpaired following a synchronous adapter raises the possibility that synchrony-induced binding 'protects' the constituent dynamic features from the aforementioned local adaptation. By this account, the reduction in performance following the asynchronous adapter is not a result of adaptation to the global asynchrony per se, but rather of the asynchrony releasing features from global processing so that feature-wise adaptation can occur.

There has been little investigation into parts and wholes with respect to dynamic high-level adaptation. Xu et al. (Xu, Dayan, Lipkin, & Qian, 2008) investigated the relationship between low-level visual components (curves) and static facial expression aftereffects (hereafter FEAE). Adaptation to curves produced repulsive aftereffects on the perceived convexity of subsequently presented curves, but also biased perception of faces' expressions when their mouths occupied the adapted position – making happy faces appear sadder following adaptation to convex curves and vice versa following concave curves. This was taken as evidence that local, low-level adaptation can have consequences on higher level compound percepts. The effect of adaptation on perceived facial expression was stronger however if the adapting curves formed the mouth within a cartoon whole-face context, indicating also a holistic aspect to the FEAE.

Going the other way, adaptation to faces only biased subsequent judgements of curves when the adapting faces were cartoons, not photographs. Adaptation to a curve matched to the photographs' mean mouth in contrast, curvature and length did bias test curve judgement, discounting these low-level explanations. Overall then, both local and holistic adaptation contribute dissociably to the static FEAE. Though each level of processing is not independent of adaptation in the other, the extent of their interaction depends on details of the adapting face context.

Xu et al. (Xu, Liu, Dayan, & Qian, 2012) again deployed curves and cartoon faces, this time dissociating the extent to which their FEAE reflects de novo high-level adaptation as opposed to low-level adaptation inherited from earlier stages of the visual hierarchy. Crowding an adapting curve (with flanking curves) reduced the effect of the aforementioned adaptation on curves but not faces, while crowding an adapting face reduced the effect of adaptation on faces but not curves. This selective reduction of the same adapter's potency for the low versus high-level stimuli at test evidences that this adaptor is acting on distinct mechanisms. The authors also found that whereas brief presentation of adapting curves reduced their adaptive effect on both curves and faces similarly, brief presentation of faces reduced their adaptive effect on faces much less than curves. This was argued to result from the relatively more sustained activation of higher levels of the visual hierarchy in response to transient stimuli (Kiebel, Daunizeau, & Friston, 2008), such that a briefly presented face nonetheless adapts face-selective areas to a relatively greater degree than lower areas. Thus, although the FEAE consists partially of lower-level adaptation to components, further higher-level adaptation to whole faces is distinguishable.

Overall, Xu and colleagues provide evidence for both the separability and interaction between component and whole-face adaptation. Both contribute simultaneously to FEAEs, but which predominates the measured perceptual outcome depends on details of experimental design and whole-face context of stimuli. With respect to our Experiment 6, it is therefore plausible that the dynamic whole-face context of relative feature timing could mediate adaptation to the single feature whose discriminability our task measures.

Alternatively, the involvement of attractive perceptual models remains a possibility. By this account (introduced in Chapter 3 above), if reduced performance for synchronous test faces reflects holistic models that prefer such temporal relationships, then the effect of adaptation may be better explained as recalibration. Whereas prolonged exposure to asynchrony might shift these models toward the adapter, no such shift would be expected following exposure to synchrony, given this is where the models are initially centred. Our final experiment therefore sought direct evidence for the recalibration of synchrony-preferring perceptual models under adaptation to asynchrony.

4.3. Experiment 7: Little evidence for adaptive recalibration of internal models of facial timing centred on synchrony

4.3.1. Introduction

In the previous section, we described two potential explanations of our timing-dependency's behaviour under adaptation: protection of individual features by holistic binding, and recalibration of predictive models centred on synchrony. While the former is novel, the latter (suggested by Cook et al., 2015) has precedent in the literature, and is assessed by our final experiment.

Predictive coding casts perception as recurrent error-correction, whereby efficient representations become honed via the interplay of feedback predictions and feedforward error signals (Rao & Ballard, 1999). That is, the function of information travelling upward through the visual hierarchy is to signal discrepancy between input and learned statistical regularities imposed downward from higher levels. In this way, the range of potential solutions to the otherwise ill-posed problem of source determination from ambiguous sensory input is constrained by prior experience, while activity is optimised by reducing redundancy.

By this account, internal representations achieve maximal correspondence with external stimuli (efficient coding, Barlow, 1961, as cited and reviewed by Friston, 2010) by taking the form of forward predictions. These predictions may be implemented as early as the retina (Hosoya, Baccus, & Meister, 2005), and are likely present throughout the visual hierarchy (Nicholas Furl et al., 2010). The 'filling in' of illusory motion trajectories has been reported in V1 (Muckli, Kohler, Kriegeskorte, & Singer, 2005), and is associated with feedback connectivity from hMT+/V5 (Sterzer, Haynes, & Rees, 2006). Optic flow elicited by biological motion specifically may be predicted in STS (Giese & Poggio, 2003), and the mirror neuron system may deploy (inversion of) generative models and an observer's motor control system to infer the intentions behind observed actions (Kilner, Friston, & Frith, 2007).

Facial expression perception has been cast in terms of predictive coding. Furl et al. (2010) showed participants predictable dynamic transitions between fearful and neutral expressions while undergoing MEG, and compared against activity elicited by frame-scrambled and static stimuli. They found relatively greater activation for the predictable movement 165ms post-stimulus in occipital areas followed (at 237ms) by higher-level activity in the fusiform gyrus and STS, as well as premotor areas. Furthermore, the direction of the trajectory (from fearful to neutral or vice-versa) biased reported perception of a subsequently presented static neutral face, consistent with its representation by a predictive model activated by the preceding stimulus.

Reinl & Bartels (2014) also showed participants increasing/decreasing dynamic fearful expressions, but used naturalistic videos of an actor. As well as containing naturally increasing or decreasing expression intensity, the sequences were played in forward or reversed order in a 2 × 2 fMRI design. By crossing emotion direction with movement naturalness in this way, the authors found sensitivity to emotion direction in the FFA only when sequences were played forward, and increased responsiveness in OFA and pSTS when sequences were reversed. This latter increased activation is consistent with prediction errors generated by the unnatural movement.

Adaptive recalibration plays an important role within the predictive coding framework, as the means by which recurrent sensory predictions and error signals hone representations toward greater efficiency. This convergence toward a stable point of maximal mutual information is effectively modelled by attractor dynamics, evidence for which has been found in the brain and IT specifically (Akrami, Liu, Treves, & Jagadeesh, 2009; Friston & Kiebel, 2009). Attractor fields have also been invoked to model facial identity encoding and recognition (Tanaka, Giles, Kremen, & Simon, 1998). For our purposes, attractor fields model face space as a similarity space in which face representations carve out basins into which visual input is attracted during recognition (Tanaka & Corneille, 2007). Thus, a face is categorised as a given identity if it matches or is sufficiently similar to its internal representation, relative to others in its neighbourhood.

The internal models of facial change posited by Cook et al. (2015) invoke similar attractor dynamics, but such theories also predict plastic optimisation of these internal models. Similarly to lower-level visual attributes such as colour, adaptive renormalisation to facial appearance may serve to promote perceptual constancy, removing uninformative aspects of the visual environment so as to enhance detection of change or novelty (Webster & MacLeod, 2011). On a computational level, the exploitation of statistical dependencies so as to reduce redundancy has been shown to be beneficial to facial recognition performance (Bartlett, 2007). Regarding timing perception, adaptive recalibration of desynchronised audio-visual events toward greater perceived synchrony has been reported (Fujisaki, Shimojo, Kashino, & Nishida, 2004), including for speech (Navarra et al., 2005).

To examine whether adaptation causes perceived feature synchrony to recalibrate (reflecting the shifting of perceptual models of facial timing toward the adapter), we measured the build-up of adaptation over time. This entailed presenting a constant asynchronous adapter for an extended period while asking participants to repeatedly judge its eyebrow-mouth relative phase. If participants' models of perceptual synchrony shift under adaptation, this should manifest as phase judgements of the asynchronous adapter becoming less asynchronous over time. Because there are two phases at which our stimulus' eyebrows and mouth can be synchronous (0° and 180°), asynchronous adapters near both of these were tested. Attraction toward synchrony then predicts perceptual shifts in opposite directions: judgments of a 225° adapter would be predicted to shift downward toward 180°, whereas judgments of a 315° adapter would be predicted to shift upward toward 360°. The discovery of such opposing effects would constitute strong evidence that facial synchrony is indeed what adaptation shifts perception toward.

4.3.2. Methods

Participants

9 participants took part in the study, 8 of whom (including both authors) who are experienced psychophysical observers. All barring the experimenters were naïve as to the purpose of the experiment.

Stimuli and Apparatus

Stimuli, progress counter, monitor resolution, and viewing conditions were as described in Experiment 4. The monitor refresh rate was increased to 120Hz, and the background colour was changed from black to the mean of the stimuli in LAB colourspace. Matching the background colour to our faces' skin tone was intended to ameliorate potential confound by low-level (colour and luminance) adaptation. No eye tracker was used so as to minimise interruptions, given we were taking measurements always at the same stimulus level over time. Participants provided their input by adjusting a dial that controlled the non-adapting face's eyebrow-mouth relative phase. They submitted each completed adjustment by pressing a key on the computer keypad.

Design and Procedure

Participants completed a total of 800 trials split over 8 blocks (undertaken at participants' convenience). Because response times varied widely across participants, the entire experiment could take anywhere from 1 hour up to 2 hours and 15 minutes to complete. Each block contained 100 trials within which all parameters were constant, i.e. adapting face's phase, adjustable face's starting phase, and positions of the two faces (these parameters were randomised between blocks). Both faces were always presented upright. The final phase of the adjustable face submitted by participants in each trial (relative to the adapting face's phase that they had to match) was taken as our dependent variable.

The adapting face's eyebrows and mouth could oscillate at either 225° or 315° of relative phase, and it did so continuously throughout each block (though its cycle reset at the beginning of each trial). The adjustable face occupied the opposite side of fixation and began each trial oscillating at ±45° relative to the adapter (e.g. 180° or 270° in the case of a 225° adapter). Neither face contained misaligned eyebrows. 100 sequential trials per adapting face phase, per adjustable face starting phase, per adapter position were presented (left or right; 100 × 2 × 2 × 2 = 800).

Within each trial, participants viewed both faces either side of fixation and had to adjust the adjustable face to match the phase of the adapting face. There was no time limit imposed to do so, and the next trial began immediately upon submission of a response. Each block therefore consisted of repeated adjustments upward/downward (depending on the starting phase of the adjustable face relative to the adapting face), with the intention being that a systematic shift in the perceived phase of the adapting face would manifest as an upward/downward trend in responses over the block. Starting phase of the adjustable face was counterbalanced to ameliorate potential confound from systematic under/overshooting, e.g. submitting once the adjustable face becomes 'close enough'.

4.3.3. Results

Each participant initially provided 8 time series (100 trials long; one per condition combination), which was coded relative to the adapting face's phase. For example, if a participant adapted to 315° of eyebrow-mouth relative phase and matched it by adjusting to (and submitting) 320°, this would be coded as 5°. If a participant adapted to 225° and made a response of 210°, this would be coded as -15°. A veridical response when matching either adapter would be coded as 0°. By predicting that adaptation should shift perception toward synchrony therefore, we are predicting a positive trend in the 315° adapt condition, versus a negative trend in the 225° condition.

Because participants responded at their own pace, overall duration of blocks varied widely (see Figure 4.4). This precludes any one-to-one correspondence in progress through blocks as measured by physical time versus number of trials completed. It also complicates summarising and interpretation in physical time, because data points do not coincide temporally across participants. The following analyses serve to test the hypothesised divergence of responses between the 225° and 315° adaptation conditions, in both a temporal and trial-wise fashion.

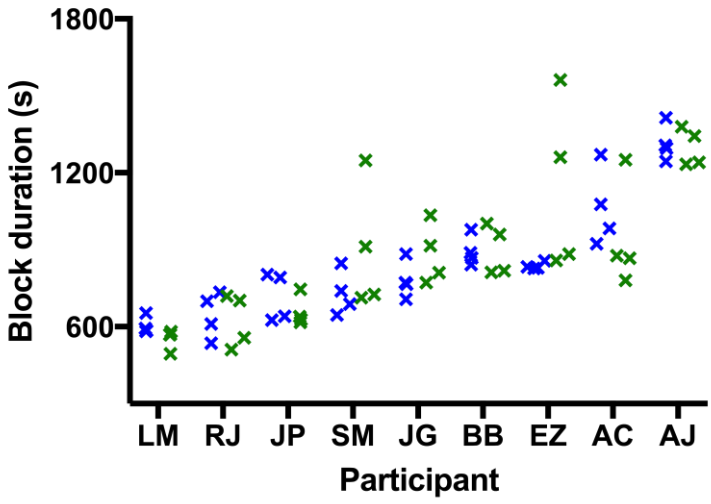


Figure 4.4. Duration of each block for each participant (ordered from fastest to slowest). Blue denotes adapting to and matching 225°, green denotes adapting to and matching 315°. Each participant completed four blocks per adapting condition, corresponding to the two adapting hemifields and two adjustable face relative starting phases (above/below the adapting phase; fully crossed).

Responses pooled over all participants, adapted hemifields and both adjustable face starting phases are plotted against time in Figure 4.5. To assess the differential build-up of adaptation, participant responses were regressed against time and adapting phase (as a binary dummy variable), including an interaction term.

As can be seen in Figure 4.5, the data was noisy relative to the size of any trends and this resulted in a model accounting for only a small proportion of variance (R^2 = .0065). Reflecting the large number of data points however, the model was also highly significant ($F(3, 7196) = 15.77$, $p = 3.23 \times 10^{-10}$). Although the regression achieved very high power when pooling over participants therefore, the presence of linear effects of adaptation over time is not supported. Furthermore, the time × adapting condition interaction was non-significant ($\beta = 0.0014$, $p = 0.23$). Divergence of responses between the two adapting conditions is therefore not supported.

We also conducted trial-wise analyses allowing us to smooth each block's time series and collapse (average) over them. Correspondence of data points across blocks and participants also enabled calculation of point-wise dispersion, and the aggregation of early versus late trials so as to directly compare adapting conditions within each interval. Our prediction of the build up of divergent adaptation effects over time amounts to predicting a difference between the two conditions in later trials that is not present in earlier trials.

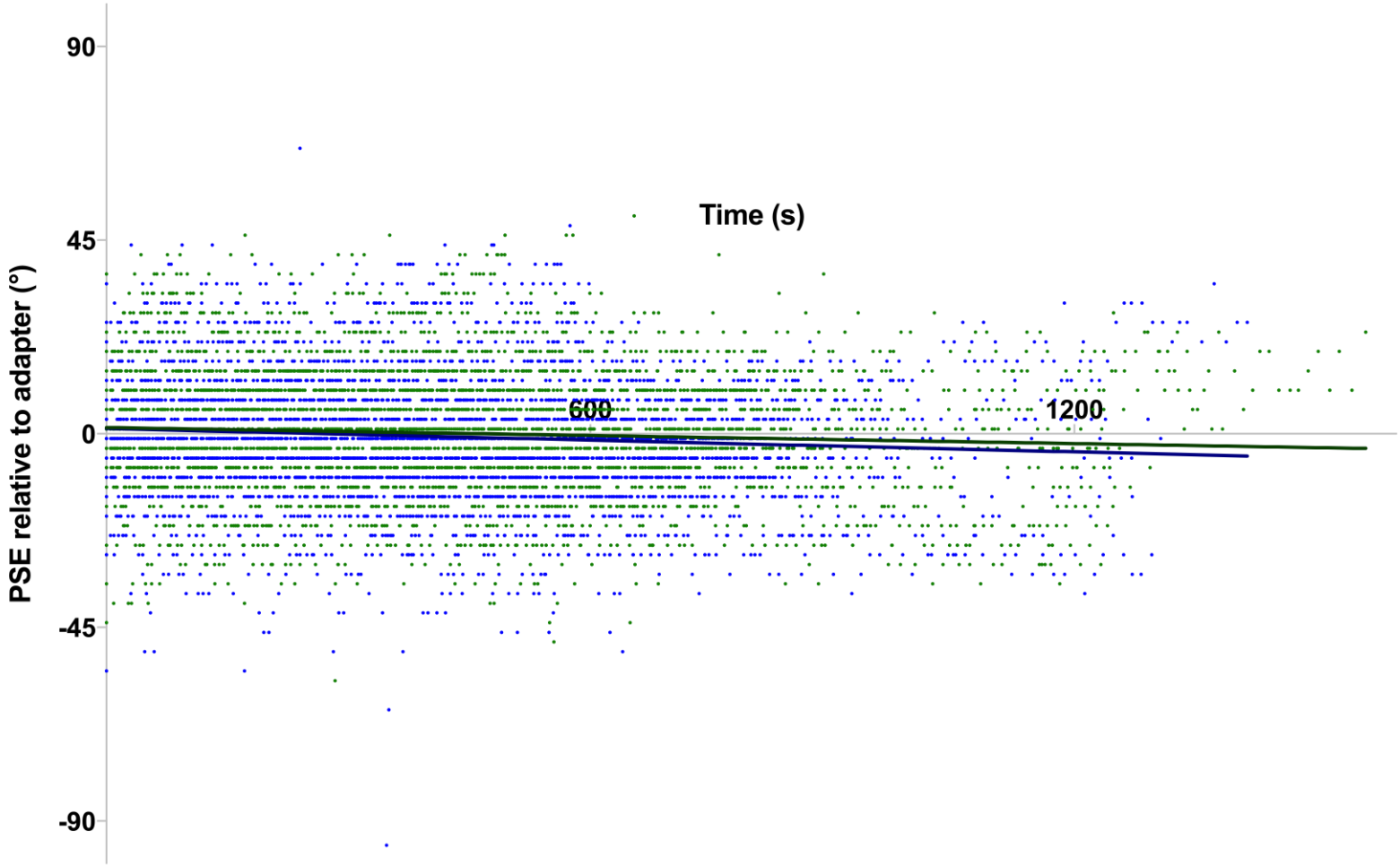


Figure 4.5. All responses from all participants plotted against time. Responses are relative to the adapting face's eyebrow-mouth phase to be matched: 225° (blue) or 315° (green). 0° therefore denotes a veridical response, and perceptual attraction toward synchrony over time would manifest as a downward trend in blue, and an upward trend in green. The lines of best fit show very little slope, however.

Figure 4.6 shows mean responses of participants plotted against trial number. Each block's time series was first smoothed by applying a 5-trial wide moving average. Collapsing of time series within each participant was achieved by averaging over all blocks for each adapting phase on a per-trial basis. Trial-wise error bars then reflect the 9 collapsed data points contributing to each final mean (one per participant).

Visual inspection of Figure 4.6 suggests a divergence in the time series roughly between trials 46 – 70. To assess this, we defined an equivalent length interval between trials 21 – 45 (inclusive) and calculated an average response for each participant and adapting condition to these early and late intervals (Figure 4.7). A 2 × 2 ANOVA revealed a highly significant interaction between interval and adapting condition only ($F(1, 8) = 11.57$, $p = .009$), reflecting that adjustments to match the 225° adapter were marginally significantly lower than to the 315° adapter only in the late interval (mean difference between adapting conditions in the early interval = 0.46°, $SD = 3.62°$; pairwise $t(8) = 0.38$, $p = .712$; versus mean difference in the late interval = -2.94°, $SD = 3.91°$; pairwise $t(8) = -2.25$, $p = .054$).

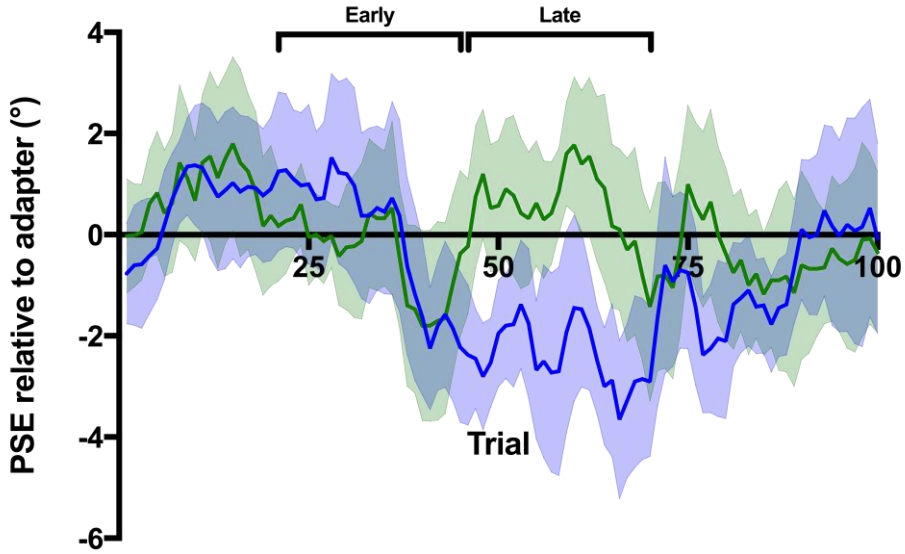


Figure 4.6. Mean collapsed smoothed responses ±1 SEM plotted against trial index. All blocks were first smoothed via moving average. We then collapsed over blocks within each adapting phase (225° in blue, and 315° in green) for each participant. This plot therefore shows the mean over participants for the two adapting phases. We also collapsed over trials in the 'Early' and 'Late' intervals, see Figure 4.7.

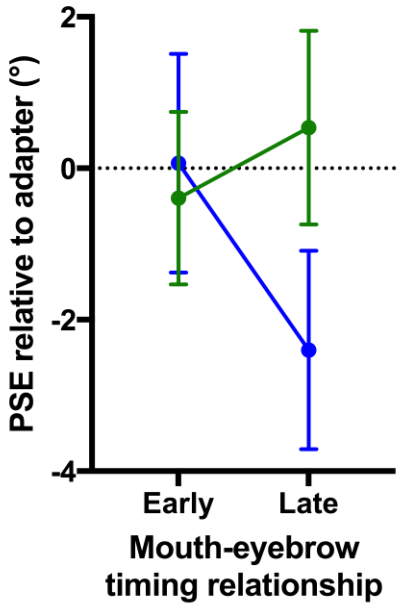


Figure 4.7. Mean response ±1 SEM across trials 21 – 45 (Early) and 46 – 70 (late), for the 225° (blue) and 315° (green) adapting conditions. This analysis follows the smoothing and collapsing detailed in Figure 4.6.

4.3.4. Discussion

Our participants' judgements of the adapting faces' eyebrow-mouth relative phases (as measured by adjustment to match) did not shift toward synchrony over time, and equivalently responses to the opposing adapting phases did not diverge over time. Some evidence of overall divergence in an interval of trials was found (which was not present in the preceding interval), but for reasons described below this can only be considered suggestive.

The strongest theoretical reason to favour the outcome of our regression in physical time over trial-wise analysis is that the former is the scale on which neural adaptation operates. Given the widely varying duration each trial represents across and within participants, ascribing an effect that is visible in trial-wise but not time-wise analyses to adaptation would imply that one's rate of adaptation tracks rate of responding in our task; there is little reason to believe such a relationship exists. On the other hand, aligning responses by trial index allows for straightforward smoothing and aggregation, increasing our power in the sense of revealing structure in the noisy raw data.

Notably however, this trial-wise structure is also not captured by linear trends. Instead, divergence between judgements of the adapting faces' phases was specific to a 25-trial interval occurring roughly half way through the mean time series; not only did this difference disappear subsequently, but there is also enough divergence in the first 5 trials or so that the early-versus-late interval comparison only reaches significance with their exclusion. Thus, although a significant divergence in judgements of facial timing following adaptation to opposing phases that is not present in preceding trials can be reported, this hinges on careful selection of the intervals being compared. While discarding of the first few trials of each block is common due to their unreliability, it is less clear why perceptual attraction toward synchrony should abate prior to removal of the adapting stimuli. The differing paces at which participants completed blocks means that later trials represent a wider range of exposure times to the adapter, obscuring their interpretation. The divergence is also smaller than would be expected for attraction. Overall, our significant interaction justifies further investigation (with enforcement of fixed-duration trials), but it must currently be regarded as exploratory.

In sum, our final experiment does not support the effect of adaptation to asynchronous global facial motion being attraction toward perceived synchrony. The alternative – that asynchronous motion allows per-feature adaptation precluded by holistic binding in the synchronous case – remains a plausible line of enquiry.

CHAPTER 5. Facial motion sensitive brain areas are modulated by relative feature timing

5.1. Introduction

In the preceding chapters, we presented several behavioural results demonstrating how the binding of dynamic facial features depends on the relative timing of their movement. We also attempted to characterise the underlying neuroscience through divided visual field and adaptation experiments. This final experimental chapter presents a more direct approach, deploying functional imaging to assess the modulation of facial motion sensitive brain areas by feature relative timing.

Following on from the general introduction, this chapter begins with a overview of the literature that is more focussed on the neuroscience of facial motion processing. In subsequent sections, we then present a two-stage fMRI experiment which both localises facial motion sensitive areas using naturalistic video stimuli, and reveals their modulation by facial timing using a phase encoded design.

The dominant models of cortical organisation for face processing propose a segregation of function in relation to static/invariant and dynamic/variable aspects of facial form, which are thought to be processed in the ventral and dorsal streams respectively (see General introduction above). Whereas the main destination of the ventral stream is FFA, the STS is the primary dorsal component of the core face network. However, breaking with this now classic view of the face network, recent imaging research suggests that input to the STS may come directly from motion area V5/MT as opposed to OFA (see below). Furthermore, several new anterior face-selective regions (forming an extended face network) have only recently been revealed by imaging with the use of dynamic localisers.

The STS has been implicated in a wide range of roles beyond facial motion processing, which may be a result of the sulcus' recruitment into different functional networks in a task-dependent manner (Hein & Knight, 2008). The extensive connectivity of the STS with many brain structures supports this view. Another, non-mutually exclusive perspective is that the multimodal and integrative properties of the STS subserve dynamic person perception (Yovel & O'Toole, 2016). This account concords with its dual facial and biological motion sensitivity (possibly in alternating regions in cortex; Weiner & Grill-Spector, 2013), combined with its role in audiovisual integration (Beauchamp, 2005). Overall, it is becoming increasingly clear that the role of STS in person and/or social perception is nuanced and entails greater complexity than is captured by the dominant core face network models.

Another deviation from the traditional models is the emerging finding that STS seems to receive input from motion area V5/MT, rather than OFA. Pitcher, Duchaine, & Walsh (2014) scanned participants using fMRI both before and after administering TMS to STS or OFA. Whereas disruption to OFA reduced the response of STS to static but not dynamic faces, disruption to STS reduced its response to dynamic, but not static faces. Whereas disruption to OFA reduced activity in FFA for both types of stimuli (supporting the expected cortical connectivity between these areas), dynamic information appears to reach the STS by a route other than OFA.

Furl, Henson, Friston, & Calder (2015) localised V5, STS, OFA, and FFA using fMRI, before applying dynamic causal modelling (hereafter DCM) to find the most likely pattern of functional connectivity between them. The best bilinear model (one that permits modulation of endogenous connectivity between regions by an experimental factor, in this case motion or facial form) included full endogenous connectivity, with motion information transmission between V5 and STS gated by facial form. A subsequent nonlinear model (one that permits modulation of connectivity between regions by a third region) implicated the OFA as the source of this V5 – STS modulation. In other words, DCM of the localised core face network plus V5 revealed STS was driven by motion information from V5, but this was gated by the presence of facial form encoded in OFA.

A further source of complication in interpreting the imaging results arises from experiments that appear to show sensitivity to facial motion in FFA, and to static faces in STS. Arsalidou, Morris, & Taylor (2011) for example conducted their own fMRI alongside a meta-analyses of prior imaging experiments, and found, in both cases, increased activity for dynamic over static facial expression stimuli across a range of face areas including fusiform gyrus. Engell & Haxby (2007) presented participants in the scanner with Ekman & Friesen's (1976) pictures of facial affect displaying various gaze directions. Despite the use of static stimuli, they reported greater activation in STS for emotive faces over neutral ones, and for averted over direct gaze.

In their review, Bernstein & Yovel (2015) note that most studies reporting FFA sensitivity to facial motion used morphed stimuli. The resulting motion was generated by warping the face in a way that unlikely to be naturalistic. They argue that increased FFA activity in these cases is a response to the changing form of the face. Conversely, increased activation of STS in response to static emotional or gaze-directed stimuli could represent the analysis of motion from form. Faces in nature typically do not maintain fixed gaze directions or expressions for extended periods, and so static stimuli exhibiting these could be interpreted as containing implied motion – momentary snapshots of a dynamic sequence. Bernstein & Yovel (2015) find these arguments consistent with their revised FFA-STS dichotomy of form versus motion. For our purposes, it is clear from this work that which brain areas are activated in a given facial motion imaging experiment depends upon the precise details of the dynamic stimuli employed.

Recently, several new anterior face sensitive cortical regions have been reported, two of which – in the anterior STS and inferior frontal gyrus (hereafter aSTS-FA and IFG-FA) – are consistently localisable in humans only with dynamic stimuli. The aSTS-FA was initially implicated in face processing in macaques, with Perrett et al. (1985) reporting this area's sensitivity to eye gaze and head view using microelectrode recording. Also using microelectrode recording, Eifuku, De Souza, Ryoi, Nishijo, & Ono (2004) reported a face space in macaque aSTS-FA encoding head view, before going on to segregate aSTS-FA into a caudal half invariant to symmetric views and a rostral half more viewpoint dependent and more sensitive to gaze direction (De Souza, Eifuku, Tamura, Nishijo, & Ono, 2005).

In humans, Calder et al. (2007) used fMRI adaptation to measure sensitivity of aSTS-FA to gaze direction. Prolonged exposure to leftward gaze reduced the relative BOLD response of aSTS-FA to this gaze direction (and likewise for a rightward gaze adapter). The adaptation also caused averted gaze to appear more direct when measured behaviourally. Overall then, this region appears to code the direction of gaze (or direction of attention, Carlin, Calder, Kriegeskorte, Nili, & Rowe, 2011) of others. The above studies all targeted aSTS-FA a-priori however. In reviews, Duchaine & Yovel (2015) and Bernstein & Yovel (2015) note this region has rarely been localised without use of dynamic stimuli. Based on the rationale that dynamic localisers are more naturalistic, Fox, Iaria, & Barton (2009) reported much greater success in localising the extended face network using these in comparison with static localisers. Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher (2011) reported that the aSTS-FA was responsive only to dynamic faces.

Similarly, the IFG-FA was first implicated in face processing in macaques (O Scalaidhe, 1997; Tsao, Schweers, Moeller, & Freiwald, 2008), and is better localised by dynamic stimuli (Fox et al., 2009; Pitcher, Dilks, et al., 2011). Chan & Downing (2011) localised FFA, and IFG-FA, and compared their sensitivity to whole faces, eyes alone, and faces without eyes (among other non-face stimuli). They also localised the frontal eye fields (hereafter FEF) in a subset of participants. The authors found that FFA was most responsive to whole faces, with its responsiveness to these not significantly different from eyes-masked faces, and its responsiveness to eyes alone not significantly different from the (non-face) flowers control. By contrast, they found that IFG-FA was most responsive to eyes alone, with its responsiveness to eyes alone not significantly different from whole faces but significantly greater than eyes-masked faces. In other words, whereas the FFA appears sensitive to the overall face rather than eyes specifically, the reverse appears to be the case for the IFG-FA. This region did not overlap the FEF, though the two regions were sufficiently close that the authors did not rule out the former's possible involvement in generation of eye movements.

These regions may be connected in a dorsal facial motion processing stream that extends through STS into prefrontal cortex. Ethofer, Gschwind, & Vuilleumier (2011) presented participants with faces showing static and dynamic direct and averted gaze face stimuli, combining fMRI with diffusion tensor imaging (DTI). They did not find direct fibre connections between STS and fusiform gyrus (supporting the segregation of these regions into distinct processing streams) but did find connections between right STS and the right anterior insular whose functional connectivity was increased by social attention. The left STS was found to project fibre connections terminating near left IFG. Gschwind, Pourtois, Schwartz, Van De Ville, & Vuilleumier (2012) also did not find direct anatomical connectivity between STS and OFA or FFA, but again reported projections from STS forward into lateral prefrontal areas.

Overall then, recent evidence diverges from the traditional core face network models in a few respects. Firstly, it is likely that the motion and form pathways for face processing are more segregated than has been previously suggested, both beginning in distinct occipitotemporal areas (V5/MT versus OFA) and terminating in distinct frontal regions (e.g. IFG-FA; we have not considered the form pathway here but see Chan, 2013 for a review). There are also several anterior regions comprising an 'extended' network which may only have been revealed with use of more naturalistic, dynamic face stimuli. A far wider array of cortical regions (in more anterior temporal cortex and beyond) have in fact displayed face sensitivity dependent on the stimuli used (Chan, 2013; Collins & Olson, 2014).

Our final experiment deployed naturalistic dynamic face videos to localise as many facial motion selective regions as possible, for comparison with a whole-brain analysis using phase-encoded stimuli derived from our behavioural experiments. The phase-encoded stimuli allowed us to measure the extent of each voxel's modulation by eyebrow-mouth relative timing, and we compared the extent of this modulation for regions whose activity is sensitive to facial motion versus not.

Phase-encoded fMRI designs have their origins in retinotopic mapping of early visual cortex (Sereno et al., 1995), and retinotopy remains their primary use. Stimuli typically consist of rotating chequered wedges cycling slowly around fixation, such that different sections of the visual field are repeatedly stimulated at different points in the cycle. Brain areas or voxels whose activity oscillates at the same frequency as the wedge's rotation are therefore implicated in representing a specific region of visual space, and one can ascertain which region that is by the phase of the oscillation. In this way, the retinotopy of areas such as V1 and V2 can be mapped, as can the boundary between them due to their respective maps being mirror reversals of each other. Higher visual cortical areas also display retinotopy and can be similarly mapped with inclusion of stimuli for which they are specialised (e.g. faces, scenes, motion) so long as these stimuli are spatially localised and track a cyclical path through the visual field (e.g. Hagler & Sereno, 2006; Huang & Sereno, 2013).

The principle of phase-encoding can be extended beyond spatial position to any stimulus dimension amenable to periodic modulation. The neural representation of sound frequency, somatosensory location and visual orientation have been elucidated in this way (Engel, 2012 provide a review). In all cases, voxels the activity of which oscillates with the same frequency as the stimulus may be regarded as sensitive to that stimulus' value. The stimulus developed for our behavioural experiments exhibits periodic facial feature motion, and so is naturally suited to such an analysis. We therefore presented dynamic faces whose feature timing cycled through all the possible relative phases at a fixed rate, and measured the extent to which each voxel's activity oscillated at that rate. This was repeated with the stimulus oscillating at two different rates, the rationale being that voxels sensitive to relative feature timing should show periodic activation in both cases.

5.2. Methods

5.2.1. Participants

Nine healthy participants took part in the study, all of whom were right handed. All reported normal or corrected to normal vision. All were naïve as to the purpose of the experiment.

5.2.2. Scanning apparatus and parameters

Participants were scanned with a 32 channel head coil in a 3-T Phillips Achieva MRI scanner at the Sir Peter Mansfield Imaging Centre at the University of Nottingham. Functional images were acquired in multiple scans over one to three sessions (see below for scan durations) using 2D gradient echo and SENSE with a reduction factor of 2. The following scan parameters were used: flip angle = 75°; TE = 35ms; TR = 2000ms; dynamic scan time 2s; voxel size = 3mm isotropic; number of axial slices = 34; matrix 64 × 64. Alignment scans were collected at the beginning of each session with the same block centre and orientation as that session's functional images. These were used to align the functional data with structural scans which were available for all participants from previous scanning sessions.

Stimuli were presented on a 24in Cambridge Research Systems BOLDscreen positioned behind the scanner bore, with a resolution of 1440 × 1080 and a refresh rate of 100Hz. Participants viewed the display via a mirror 17cm away from their eyes, which was 102cm away from the display for a total viewing distance of 119cm. As in the behavioural experiments, stimuli were presented using MATLAB and Psychtoolbox. For the sham task during the phase-encoded blocks, participants responded using a fibre-optic button pad with their right index and middle fingers.

5.2.3. Dynamic localiser

Stimuli

To localise brain regions selective for facial motion, the BOLD response to short clips of naturalistic facial movement were contrasted with the response to static frames from the same clips, and to spatially scrambled versions of those clips (see supplemental videos).

Each clip was 2 seconds in length, and consisted of a head and shoulders portrait view of a speaker. They were recorded in previous work (unpublished) in which the participant engaged in a challenging verbal task to camera, requiring the participants to repeat simple question and answer jokes many times over. The camera was operated by an experimenter who remained out of shot. The raw footage contained a mixture of facial speech with frequent errors and concomitant emotional expressions, conversational expressions, and social eye and head movements. We favoured these spontaneous moments when selecting the clips for the localiser, so that a large proportion of the facial motion included was not posed but instead represented a broad range of incidentally captured naturalistic behaviour. Actors were recorded in front of a green screen which we replaced with neutral grey for presentation in the scanner.

For the static face condition, individual frames were selected at random from the clips and presented for 2 seconds.

For the scrambled face condition, the clips were spatially subdivided into a 12 × 16 grid (squares being roughly the width of an eye) and randomly rotated each by 0°, 90°, or 270°. This had the effect of breaking up the coherence of the facial motion, while preserving overall motion energy, luminance, colour, and most spatial frequency and phase information (contours were necessarily introduced at the boundaries of the squares).

In all cases, the faces subtended about 8 DVA and were presented centrally. A fixation dot was placed in the centre of the screen. The result of this was that participants mainly fixated a region between the eyes (though this was not always the case given some occasional head movement).

Design & Procedure

The presentation cycled through 8 dynamic clips, 8 scrambled clips, 8 static frames, and 16 seconds of blank (mid-grey) screen 5 times per run. Each run therefore took a total of 320 seconds, and participants completed a total of 4 runs. The ordering of the dynamic localiser and phase-encoded runs was balanced across each participant's sessions so as to avoid fatigue systematically impacting results, and brief periods between runs allowed for resting. There was no sham task in the dynamic localiser; participants simply had to fixate.

5.2.4. Phase-encoded design

Stimuli

To measure the degree to which a brain region modulates with feature relative timing, we presented an animated avatar whose eyebrows and mouth were moving periodically at 1.5Hz (see general methods for our behavioural experiments above). The relative timing of the eyebrows and mouth were varied smoothly such that the full range of relative phases was presented over the course of 24 seconds. Brain activity that modulates at 1/24Hz is then indicative of sensitivity to relative feature timing. This procedure was repeated with the feature relative phase cycling over 32 seconds, in search of voxels the activity of which modulated at 1/32Hz. The gradual cycling of relative feature phase was achieved by making the oscillation of the mouth slightly faster than the eyebrows (by 1/24Hz or 1/32Hz).

The face subtended 10 DVA and was presented centrally. A fixation dot was placed equidistant between the eyes and the mouth, thereby occupying a position around three-quarters of the way down the bridge of the nose.

Design & Procedure

The stimulus presentation cycled through the full range of relative feature phases 8 times per run. Each run of the 24 second cycle stimuli therefore took 192 seconds, while the 32 second cycle stimuli took 256 seconds. Participants completed 4 runs of each rate. To maintain attention on both the eyebrow and mouth regions, a sham task was included whereby a small blue dot would appear for 100ms either between the eyebrows or in the centre of the mouth region at random times (once every 8 seconds on average). Participants had to indicate using the button box when they saw the blue dot appear either above or below fixation. Responses were monitored during scanning to ensure attentiveness, but were not recorded or analysed.

5.3. Results

The cortical reconstruction for visualisation of results shown in the figures below utilised the participants' pre-existing structural scans and the FreeSurfer software package. All analyses of functional data were conducted in the aligned voxel space however, not on the cortical projections. Summary statistics were obtained separately for each participant and compared using t-tests. In other words, we did not average functional data across participants or seek to reconcile them to a common spatial frame. This avoided the many pitfalls and trade-offs inherent in matching brains of different shapes (Klein et al., 2009; Yassa & Stark, 2009).

Alignment and analyses were conducted using the mrTools toolbox for MATLAB. The initial inplane scan for each session was aligned to the participant's structural scan, and the functional scans were subsequently aligned to the inplane. Motion correction was then applied to each functional scan, to minimise the impact of minor head movements during the session. These alignment steps were achieved using Robust Intensity-Based Alignment (Nestares & Heeger, 2000), with the end result that all scans from all sessions were aligned to a common frame according to each participant's structural scan.

The four functional image time series for each design (dynamic localiser; phase-encoded with 24s cycle; phase-encoded with 32s cycle) were then spatially smoothed using a Gaussian with standard deviation = 1.9mm before being averaged together, yielding three averaged time series per participant for analysis. All subsequently described analyses were applied to each participant's averaged functional data individually.

For the dynamic localiser, a standard General Linear Model analysis with separate regressors for the three experimental conditions (dynamic, scrambled, and static faces) was conducted. The design matrix was convolved with a double Gamma hemodynamic response function with the following parameters: shape of the positive Gamma = 6; shape of the negative Gamma = 16; scaling ratio = 6. Two voxelwise t-tests for each of the contrasts (dynamic faces versus scrambled faces) and (dynamic faces versus static faces) were then conducted. The resulting voxelwise *p* values were adjusted for multiple comparisons using false discovery rate by the adaptive step-up method. We considered voxels sensitive to dynamic facial motion if both of these comparisons were significant at the corrected $p < .05$ level with effects in the same direction.

Figure 5.1 shows the results of this analysis, for each participant and inflated hemisphere. Large swathes of occipital, temporal, and frontal regions can be seen to have been localised, but with a high degree of variability in overall activation across participants.

For the phase-encoded designs, the time series were detrended using a high-pass filter and spatially normalised by dividing by the mean signal, before applying Fourier analysis to calculate the correlation of each voxel's fMRI signal with a sinusoid describing the cycling of the eyebrow-mouth relative timing. Correlation is defined as a signal-to-noise-ratio, where the signal is the amplitude in Fourier space at our stimulus' frequency (1/24Hz or 1/32Hz), and the noise is the sum of amplitudes at all other frequencies. A phase-encoding correlation of 1 would therefore denote a voxel whose signal traces a perfect sinusoid of the same frequency as our stimulus.

Figure 5.2 shows the results of this analysis, for each participant, stimulus frequency, and inflated hemisphere. Again there appears to be wide variability in the amount of modulation by facial relative feature timing across participants.

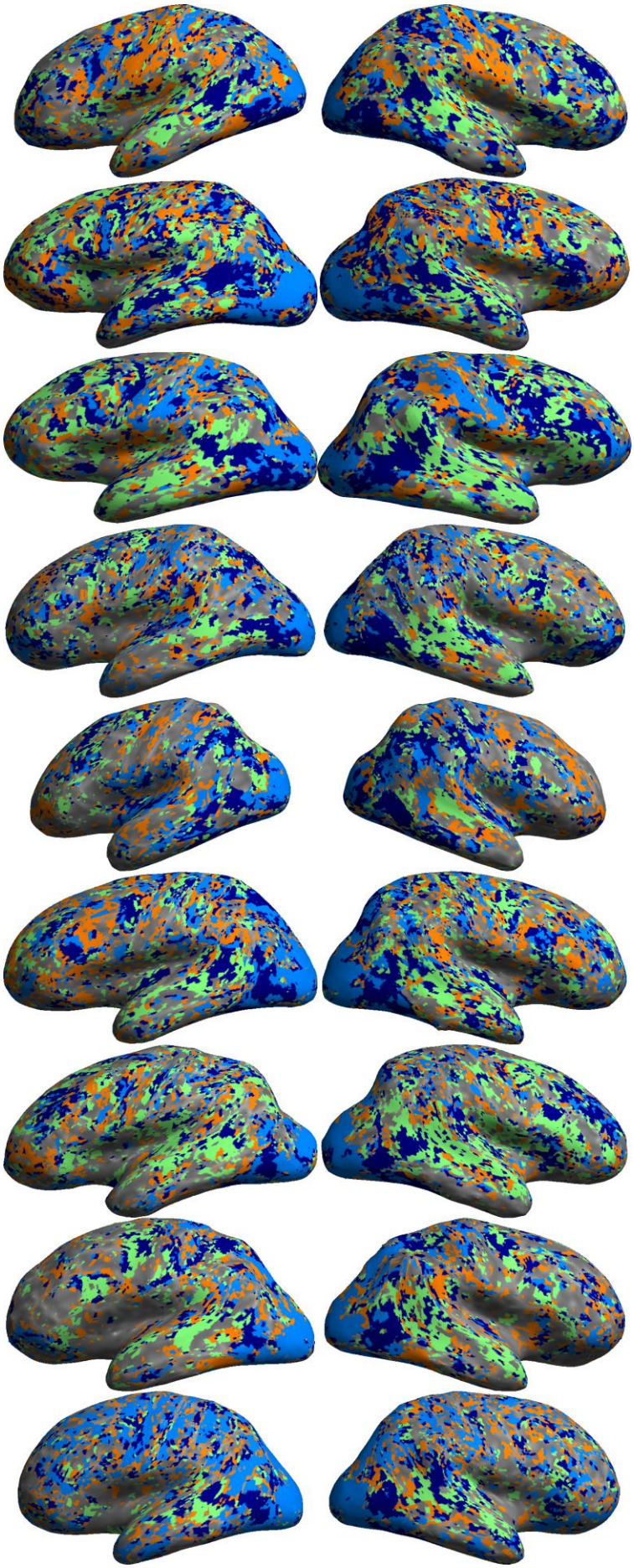


Figure 5.1. Per-participant cortical surface projections of the two dynamic localiser contrasts and their intersection. Green denotes the dynamic – scrambled contrast where FDR-adjusted $p < .05$. Orange denotes the dynamic – static contrast where $p < .05$. Blue denotes both contrasts significant. Dark blue denotes both contrasts significant with effects (β values) in the same direction. These dark blue regions were considered facial motion sensitive for the purposes of comparing modulation by relative feature timing (see Figure 5.2).

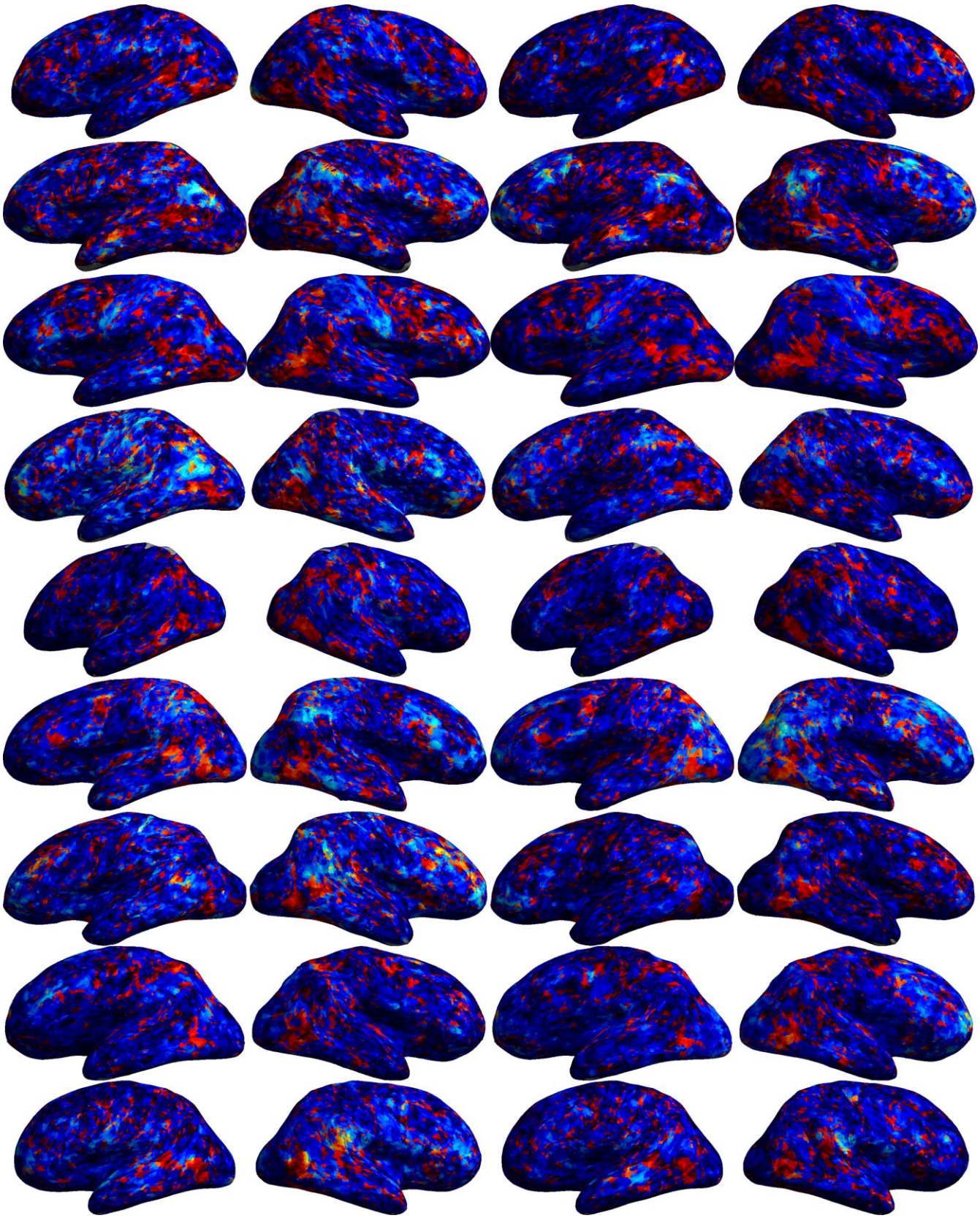


Figure 5.2. Per-participant (rows) cortical surface projections of the correlation of fMRI signal with our phase-encoded stimuli, at 24Hz (left) and 32Hz (right). Brighter colour denotes greater modulation of activity at the frequency of our stimulus. Red denotes brain regions considered facial motion sensitive for comparison against blue brain regions (the former corresponds to the dark blue colouring of Figure 5.1).

To assess the extent to which brain areas sensitive to facial motion show greater modulation by relative feature timing, we used the localiser to subdivide each participant's brain into facial motion-sensitive voxels versus not, and calculated the mean phase-encoding correlation for each. This was done separately for the 1/24Hz and 1/32Hz stimulus frequencies. Voxels outside the brain were excluded by thresholding the mean fMRI signal of the averaged phase-encoded time series; only voxels whose mean was above 200 were included. This fixed threshold was possible because of the large difference in signal between the two spatial areas, producing a clear and easily partitioned bimodal distribution that was similar across participants and scans (Figure 5.4 and Figure 5.5).

Mean BOLD modulation for the 1/24Hz stimuli was significantly greater within facial motion sensitive areas versus outside those areas (mean difference = 0.012, *SD* = 0.014; pairwise $t(8) = 2.72$, two-tailed $p = .03$). Similarly, BOLD modulation for the 1/32Hz stimuli was significantly greater within motion sensitive brain areas versus outside (mean difference = 0.012, *SD* = 0.012; $t(8) = 2.87$, $p = .02$; see Figure 5.3).

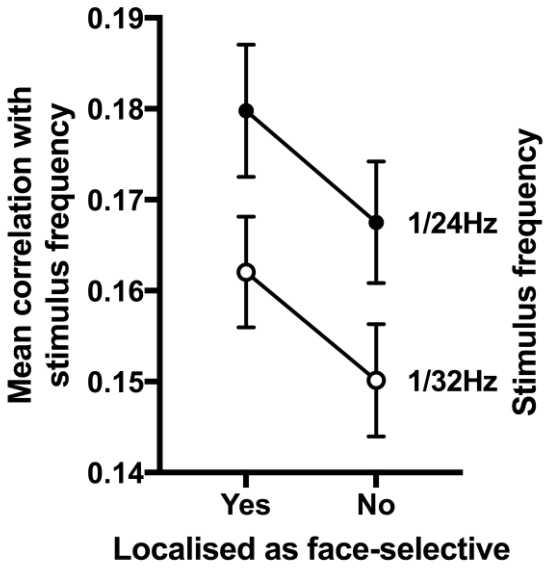


Figure 5.3. Mean modulation of fMRI signal at the stimulus frequency, for participants' brain areas identified by the dynamic localiser as face selective versus not. Error bars are ±1 SEM with n = 9, i.e. each participant contributed one mean phase-encoding correlation per point.

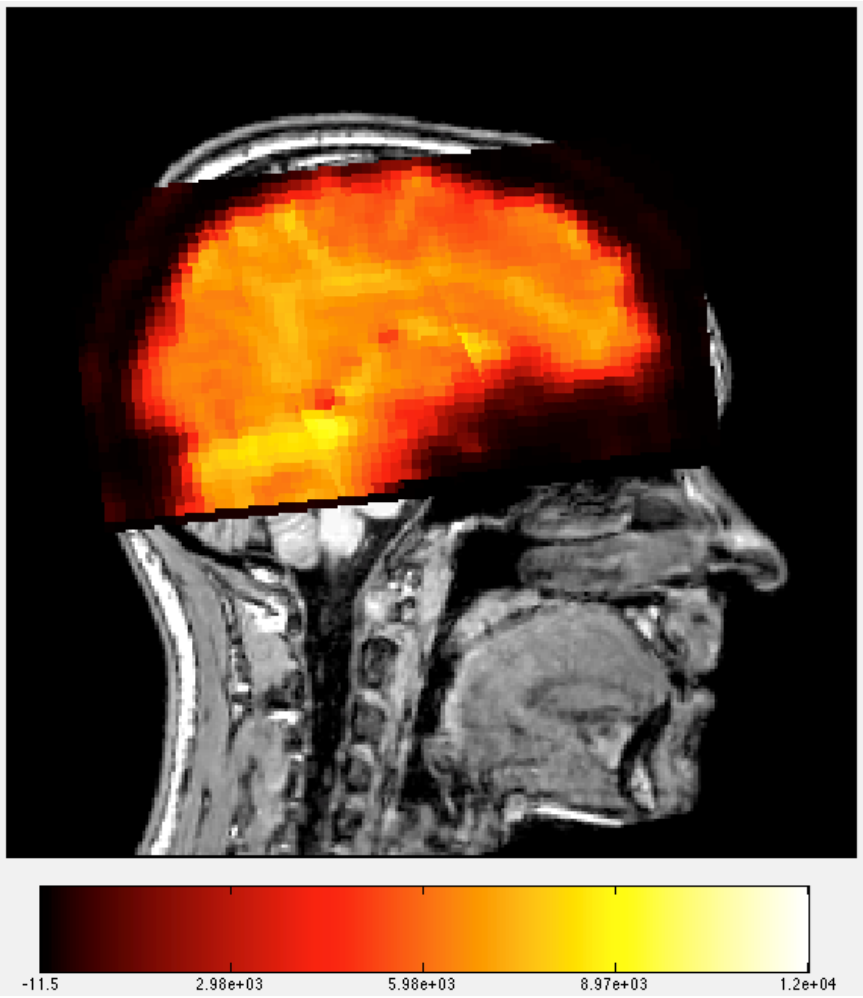


Figure 5.4. Mean signal of an illustrative averaged functional time series for one participant. The mean signal outside the brain (i.e. surrounding tissue, bone, and air) is much lower than within the brain.

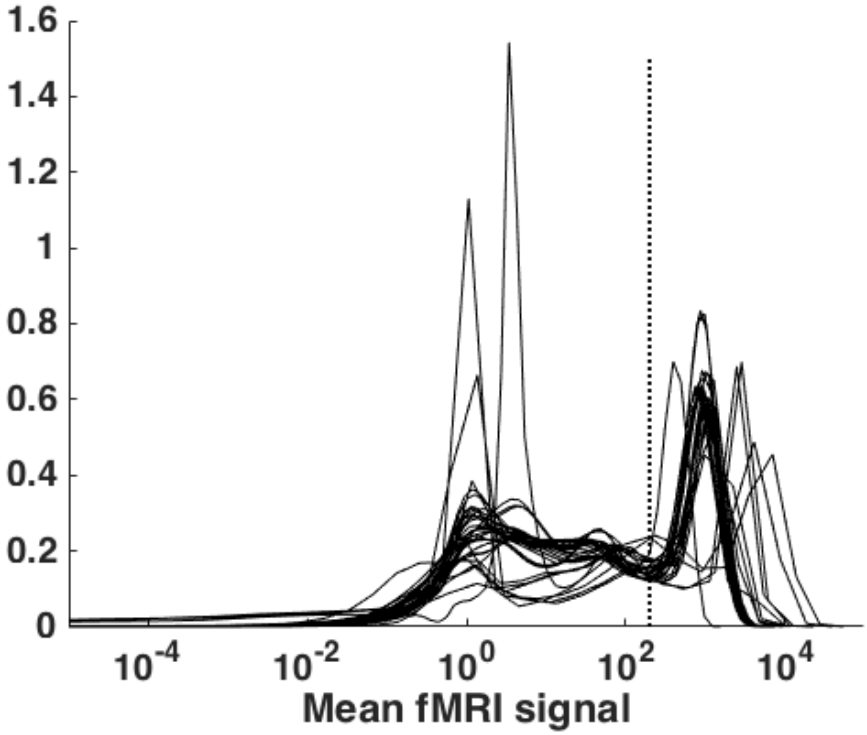


Figure 5.5. Density estimates of mean fMRI signal for all averaged functional time series, for all participants. Voxels with high mean signal are likely to lie within the brain, whereas those with low mean signal are not. Voxels of all functional time series clearly subdivide into similar bimodal distributions. All voxels below 200 (dotted line) were considered non-brain and excluded from analysis.

5.4. Discussion

Using a simple and anatomically agnostic analysis, we have demonstrated that brain regions sensitive to facial motion are more modulated by the relative timing of facial features. Because of the dual contrasts used for localisation, we can be confident that these regions are sensitive to facial motion. Because our phase-encoded stimuli varied only in terms of eyebrow-mouth relative timing low-level temporal and spatial properties were held constant e.g. the same area of visual field was stimulated and there was no overall variation in average motion speed. We can therefore exclude these variables as contributing to the observed differences in modulation.

Temporal regularities are believed to be an important cue by which visual cortex learns how features group to form coherent objects. The problem of object recognition may be cast as the discovery of invariances – derivable features that remain constant across wide variation in an object's visual appearance (Wallis & Bülthoff, 1999). If two very different visual inputs are generated by the same object (dependent on pose or viewpoint, for example), how can the brain know that these visual inputs share the same physical origin?

One clue is that visual inputs from the same object are likely to be contiguous in time. Thus, a neural system may judge that several differently appearing viewpoints do in fact belong to the same object if transformation between them occurs smoothly and consecutively. Biologically plausible neural networks have demonstrated how such a learning rule may make effective use of temporal statistics for learning invariance (the trace rule, Wallis & Baddeley, 1997). Miyashita, (1988) trained macaques with random fractal patterns in a working memory task, and subsequently recorded from pattern-sensitive neurons in anterior temporal cortex. They found that the subset of patterns to which a given neuron responded did not share geometric features, but rather they tended to be patterns that were presented close to each other in time during the working memory task. In other words, the recorded neurons appear to have learned temporal conjunctions rather than similarity of appearance.

It is possible to 'break' this invariance-from-temporal-coherence mechanism experimentally, by violating the assumption of stable object identity. Wallis & Bülthoff (2001) trained participants on CGI face models that rotated (thereby presenting a smooth range of viewpoints). The faces were subdivided into groups, and participants were asked to discriminate between pairs of faces taken either from the same group or different groups. Crucially, during rotation the faces morphed between members of their group, thereby violating the natural stability of facial identity across viewpoint. At test, participants were worse at discriminating faces from the same group, indicating that they tended to see faces from the same group as having the same identity. No such confusion occurred when the temporal coherence of the rotation was destroyed by randomising the frame order during training however, indicating that the contiguous geometric transformation in time was necessary for the erroneous assimilation of identities.

Temporal regularity therefore plays a fundamental role in the recovery of environmental structure from visual input. The trace rule has also been successfully applied to the learning of biological motion trajectories (Sigala, Serre, Poggio, & Giese, 2005). Overall, there is a large extant biological and computational framework within which the development of cross-feature timing priors is plausible. Though much of it focuses on object recognition, the theory is readily extendable to the learning of temporal patterns and natural motion sequences.

Our question was answerable without recourse to neuroanatomy in the above analysis, but two further lines of study are possible. Firstly, the phase aspect of our phase-encoded data has not been considered. We have shown that some brain areas are sensitive to facial timing but not which timings. Secondly, though the inflated cortical projections are presented in Figure 5.1 for visualisation, no claims regarding which brain areas are implicated have been made.

No retinotopy-like cortical map was obvious to visual inspection, and the phase data showed as much variability across participants and stimulus rates as is visible in Figure 5.1 and Figure 5.2. Given the small spatial extent that cortical face patches may occupy and their close proximity with non-face sensitive areas (Weiner & Grill-spector, 2012; Weiner & Grill-Spector, 2013), it may be that higher resolution fMRI would be required to reveal a relative feature timing map. For comparison, orientation pinwheels occupy just under under $1mm^2$ on average (Bonhoeffer & Grinvald, 1991).

The finding that at least some facial motion sensitive regions prefer some dynamic feature relationships is novel. This initial whole-brain analysis therefore motivates potential future investigation into which face processing areas have such a preference, and whether a single relative timing (such as synchrony) is preferred by all or if there is differentiation.

CHAPTER 6. Conclusion

6.1. Overview

This final chapter first reviews the previous experiments and their main conclusions. It then places these results in the broader literature context, including how they inform and advance current knowledge. Some limitations of the research are discussed, and future directions highlighted. The chapter then closes with a brief overall summary.

6.2. Summary of results

Experiment 1 extended Cook et al.'s (2015) illusory slowing of feature movement to an explicitly whole-face context using adaptation. This served to confirm eyebrow movement rather than eyelid blinks can serve as a dynamic facial feature, and provided further evidence that face specific processing underlies facial feature speed illusions as slowing occurred only for upright faces.

Experiments 2 and 3 extended Cook et al.'s cross-feature interactions from perception to performance. Movement of one feature not only affects subjective judgements of movement in another, but (as Experiment 2 shows) it can also impair sensitivity. This is noteworthy because unlike response bias, measurements of performance are not explicable in terms of a shift in decision criterion, so a stronger case can be made that these interdependencies have their origin in visual rather than cognitive processes. Experiment 3 fully interleaved the presence and absence of mouth movement, which abolished the effect of the mouth on the eyebrows, thus failing to replicate Experiment 2. The abolition of perceptual interaction following presentation of moving features in a way that emphasises their dissociation hints at the role of learned priors for this effect.

Experiments 4 and 5 measured cross-feature interference across the range of relative feature timings. Experiment 4 replicated and reconciled the findings of Cook et al., Iwasaki & Noguchi (2016), and our Experiment 2, while also uncovering a special role for feature synchrony. Performance for upright faces is unhindered by mouth movement that leads the eyebrows, but is reduced to that of inverted faces when mouth movement trails the eyebrows. Mouth movement that is synchronous with the eyebrows particularly interferes with discrimination of their movement. Experiment 5 localised this synchrony dependency to the left cerebral hemisphere, contrasting with the well-known right lateralisation of face perception but consistent with processing of facial speech.

Experiments 6 and 7 probed this special role of synchrony using adaptation. Surprisingly, adaptation to facial synchrony does not affect performance whereas adaptation to asynchrony reduces performance across all relative feature timings at test. The segregation of relative feature timings into discrete processing channels is therefore not supported. An adaptation effect that persists across all test relative timings could be local to the feature being judged (eyebrows), in which case its absence for synchrony could reflect protection of local features from adaptation by holistic binding. An alternative possibility – that internal models of facial timing are centred on synchrony but subject to adaptive recalibration – was not supported.

Finally, diverse regions of core and extended face sensitive cortex were localised using naturalistic dynamic face stimuli, and these regions were shown to be more modulated by relative feature timing. Previous research has shown that temporal regularities provide important cues regarding environmental structure, including biological motion trajectories, and learning of temporal coincidence has been demonstrated at the neural level. These imaging results therefore ground our psychophysical findings in the underlying neuroscience.

6.3. Contribution to knowledge and theoretical implications

The present work demonstrates that holistic binding of dynamic facial features is contingent on their relative timing, and in particular the synchronicity of their movement. Thus, just as the static holistic face literature emphasises spatial homogeneity (a common layout against which all faces can be represented by their difference), so we have found evidence for the exploitation of temporal regularity across features for representing their coordinated motion.

A concerted shift to dynamic face perception research has occurred only relatively recently, with the majority of papers addressing motion's role in recognition. Fewer studies have considered expressive dynamics, and fewer still have strayed from the basic expressions, despite most facial movement being comprised of a stream of nuanced conversational actions. Our stimuli distilled these complex dynamics to arrive at a principled psychophysical investigation of the role of timing and coincidence. One significant impediment of extension into the temporal domain has been the relative intractability that natural temporal variation brings; by characterising elementary components of this variation however we have contributed to development of a more systematic understanding.

The wider relevance of our findings for face perception in general therefore derives from our placing of a core theoretical tenet (holistic processing) within the context of the contemporary 'dynamic revolution'. By showing that the holistic binding of dynamic features is mediated by synchrony, we have also invoked a yet more general tenet of perception science: the learning of conjunctions by temporal coincidence. Whereas contemporary dynamic face research often begins by partitioning facial movement into the actions of isolated muscle groups (e.g. Jack et al., 2014), we have demonstrated an intrinsically Gestalt aspect to how said motion is perceived, with novel consequences for visual adaptation.

Why are features bound into such a Gestalt-like representation? Firstly, it is worth noting that the eventual unity of a percept logically necessitates the combination of its elements at some point (the binding problem, Treisman, 1999). Beyond this, we can again appeal to overarching perceptual principles by returning to the themes of efficient coding and vision as an ill-posed inverse problem. What advantages does an intrinsically unitary representation afford? If a stimulus class contains much redundancy of appearance (such as the feature layout common to all faces), individual class members can be represented more efficiently by encoding them as differences from that which is common (Haberman & Whitney, 2009). This is reflected in the norm-based coding model of face recognition and has analogies to the generation of holistic eigenfaces by principal component analysis (see Introduction). As demonstrated by PCA on faces, holistic representations can arise from the exploitation of dependencies across otherwise separate elements of the stimulus (redundancy). Once learned, holistic representations can also serve as perceptual priors constraining the solution space of the problem of recognition from noisy or incomplete input (this too has analogies in the "content addressability" of PCA: a partially occluded image projected into PCA space nonetheless backprojects to a complete image, Kohonen, 1984, as cited in O'Toole, 2011).

It is reasonable to presume that these overarching perceptual principles continue to apply in the dynamic face case, and our findings support such an extension of the (primarily static) holistic face literature. That is to say, the advantages of temporal binding lie in encoding efficiency and robustness gained by exploiting dependencies in the movement of facial features. As described in Chapter 5's Discussion, temporal coincidence plays a special role as a cue to what should be bound together, so our finding that dynamic feature binding is mediated by synchrony fits with this broader context.

Because we studied elementary aspects of facial movement perception (timing), our results speak to the level of analysis 'below' that of emotional signalling. The representations that we claim to have probed represent facial actions rather than ascribed emotional states; one may consider them 'proto-expressions' that combine into expressive streams. Because temporal structure appears central to the binding of these proto-expressions, we can say that temporal information is in fact important for expression perception, contrary to the initial elusiveness of a 'motion advantage' (see Introduction). Although the motion of our stimuli is likely to be perceived as 'speech-like' rather than resembling the basic expressions, we have argued that the latter only represent a small and comparatively uncommon subset of natural expressive behaviour. It could well be that temporal structure plays a more obviously important role in perception of conversational expressions supplementing speech; how elementary proto-expressions relate to the emotional signal of the dynamic streams they compose will have to be a subject for future research however (see Limitations and future directions below).

6.4. Limitations of present work and future directions

The claim that our findings reflect face-specialised mechanisms rests on the inversion effect of the first few experiments. It is therefore worth revisiting the validity of the inversion manipulation for this purpose. To reiterate the Introduction, we preferred this manipulation due to its ubiquity in the field and comparatively minor impact on the physical appearance of the stimulus. Perhaps due in part to both of these however, inversion is not without criticism.

Also detailed in the Introduction, inversion is held to be merely an 'indirect' indicator of face-specific processing because its minimal impact on appearance means it leaves many attributes held to be important for said processing unchanged. Given the strong evidential basis for the disproportionate inversion effect for faces however, this is less a complaint regarding the suitability of inversion as a manipulation and more reflective of our incomplete understanding of how it works. 'Indirect' here refers to inversion's relationship with holistic processing being empirically, rather than logically based – that is, interactions with the 'more direct' part-whole and composite face effects are well reported despite there being no a-priori reason to regard inversion as indicative of a special processing style (see Introduction). The fact that it *is* disproportionate for faces and has been 'cross-validated' with the other two major experimental signatures of face processing gives us confidence as to its appropriateness here.

How though does inversion gate face-specific mechanisms? Incomplete understanding aside, there have been a few suggestions. We have already noted that inversion does disrupt first-order facial information (flipping the feature layout), so that any initial detection stage relying on this to initiate face-specialised processing would be defeated (Tsao & Livingstone, 2008). We have also described a contrary quantitative account (Sekuler et al., 2004) whereby processing styles for upright and inverted faces were claimed to be the same but differing in efficiency; this was refuted by Rossion (2008). Taubert, Apthorp, Aagten-Murphy, & Alais, (2011) assessed whether holistic processing played a role in face detection versus (presumably subsequent) discrimination and found similar inversion effects for both types of task. The holistic integration of features would therefore appear to be involved in both, with the authors arguing from parsimony for a single integrative process that initiates early but benefits many subsequent kinds of face judgement. Overall it is plausible that inversion defeats an initial holistic mechanism, resulting in most likely qualitatively different (and quantitatively inferior) processing.

Though use of the inversion manipulation is justified, alternatives are available and replication using these offers a valuable exercise in 'triangulation' (i.e. convergence of distinct lines of evidence, Munafò & Smith, 2018). Horizontal misalignment of face halves as deployed in studies of the composite face illusion are an obvious example (Rossion, 2013), whereby one would expect a release from synchrony-mediated binding of the eyebrows and mouth when the spatial coherence of the face is broken. Stereoscopic presentation is another possibility. Iwasaki & Noguchi (2016) also split their faces into top and bottom halves with a dividing line, but rather than misaligning the halves they varied the perceived depth of the divider using binocular disparity. When the line appeared closer to the viewer, the stimuli appeared as a complete face occluded by the line (amodal completion). When the line appeared farther from the viewer however, the stimuli appeared as two face halves separated by a gap. This was sufficient to abolish their reported dynamic mouth-eye interaction. Interestingly, Taubert & Alais (2009) have also used stereoscopic presentation to show that the composite face effect requires 'biologically plausible' face stimuli. When the top and bottom halves appeared to be different distances from the viewer (misaligned along the longitudinal axis) the composite face illusion was abolished, while slanting of the halves over an equivalent depth such that there was no step-change at their boundary restored it. Both its appearance-preserving nature (e.g. unchanged second-order relations) and implications for stimulus biological plausibility (see below) favour binocular disparity as a potentially fruitful future manipulation.

A natural tension exists between the use of highly controlled but artificial psychophysical stimuli and that which is more naturalistic. We did not aim to present participants with nameable expressions, and indeed some of the facial motion used is not easily reproduced by a real face. Rather, several of our experiments may be regarded as probing what the visual system finds natural, given it seems likely to be optimised for facial movement that is informative and commonplace. A great divide persists between the ease with which human observers interpret faces in the wild, and our understanding of (or ability to reverse engineer) this skill. It is indeed important to test perception using input that reflects typical operating conditions, but given the current intractability of this context a more fundamental analysis also has value. One natural route from the present work toward bridging this divide would be investigation of how elemental temporal units compose into more complex trajectories. An elegant, quantitative description of natural facial variation over longer timescales (seconds) would place many currently unfeasible research questions within reach.

The preceding sections placed our findings in the context of holistic processing and expression perception, however both are likely to be culture or race dependent. Distinct facial information sampling strategies between Western Caucasian and East Asian (hereafter WC and EA) observers are well reported, as are biases and deficits in recognising and interpreting faces whose race differs from one's own. Another opportunity for future investigation therefore lies in the application of our paradigm to these questions. We consider each below.

Much of the evidence for distinct visual sampling strategies derives from eye tracking experiments, with the major result being that WCs make more fixational eye movements centred on the eyes (and mouth), whereas EAs maintain their gaze more centrally on the nose (Blais, Jack, Scheepers, Fiset, & Caldara, 2008). These distinct 'local' (high acuity within small sequential windows) versus 'global' (lower acuity window covering a wider area) sampling strategies have been suggested to reflect 'analytical' versus 'holistic' perceptual styles, or even the individualistic versus collectivistic societies of the West and East (Caldara, Zhou, & Miellet, 2010). Whatever the origin, distinct sampling strategies raise the possibility of differences in the perceptual binding of features, yet we are not aware of any research directly comparing this across observers from different cultures. It could be that the more 'global' facial information sampling of EAs could render them more susceptible to our synchrony-mediated binding effect. Dynamic feature binding holds particular promise given reports that the temporal dynamics held to be indicative of the basic expressions also differ between WCs and EAs (Jack, Garrod, Yu, Caldara, & Schyns, 2012).

One potential criticism of the present work is the use of the same male WC face throughout, whereas it is known that observers perform differently when judging faces of a race other than their own (the other race effect, hereafter ORE, reviewed by Hancock & Rhodes, 2008). Recognition memory is worse for other races (Sporer, 2001), though less so with greater exposure to the race being presented supporting the primary 'contact hypothesis' (N Furl, Phillips, & O'Toole, 2002). Notably, with respect to the ORE 'configural' and 'holistic' coding appear sharply distinguished so that whereas other race faces are processed equally holistically, configural (and featural) information from them is extracted less efficiently (Hayward, Crookes, & Rhodes, 2013; Zhao, Hayward, & Bülthoff, 2014). Expression perception has also been reported to depend on the race of the face to be judged, with highly prejudiced white observers more likely to categorise ambiguous race faces as black when displaying an angry expression, and more ready to categorise unambiguous black faces as angry (Hugenberg & Bodenhausen, 2003, 2004; Kubota & Ito, 2007 provide more examples). Thus it would be fruitful to explore the persistence (or not) of our synchrony-mediated binding effect using head models of different races, though mixed findings and a lack of dynamic expressive stimuli within the existing other-race face literature would render this exploratory rather than hypothesis-driven initially.

Though we argued above that the artificiality of our stimuli is a tradeoff for psychophysical control, there exists in principle a method of retaining these benefits while animating a photorealistic face. We have already mentioned that PCA spaces behave as "content addressable memory" because PCA spaces can only represent whatever is projected into them according to what they were trained on. A PCA space trained only on complete faces will produce representations of partially occluded faces that are nonetheless complete. Therefore, it could be possible to train such a space on the dynamic face of one individual, project another individual's facial movement into it, and backproject the result. Although the details of when and how to combine datasets and PCA spaces would need to be investigated, if it is possible to animate one face with the movement of another then this opens up the possibility of using our controllable head model as 'driving input' for the backprojected photorealistic sequence. These photorealistic sequences could then be deployed to replicate the present experiments.

We did successfully employ naturalistic stimuli for our fMRI localiser. The simplicity of the overall analysis does render it quite coarse however, with no claims made regarding cortical loci. While greater modulation by relative feature timing within facial motion sensitive areas was found, both timing-modulated and face-sensitive brain areas were widely and heterogeneously distributed. Future research could investigate which face-sensitive areas are feature timing modulated, to what extent, and whether this reflects sensitivity to one specific timing or many.

6.5. Conclusion

The present work derived from reports of dynamic cross-feature interactions, which were extended to perceptual performance and elaborated. In so doing, these results show that holistic binding of facial motion depends on feature timing and particularly synchronicity. The lateralisation of this binding suggests a role in processing of facial speech, while its behaviour under adaptation argues against typical encoding models in favour of a potential interplay between adaptation to parts and wholes. These behavioural results were grounded in neuroscience with the demonstration of timing-modulated activity in facial-motion sensitive brain areas. Future research could investigate how elemental aspects of facial motion such as timing assemble into longer trajectories, and clarify the specific timing preferences of face-specialised cortex.

REFERENCES

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

Akrami, A., Liu, Y., Treves, A., & Jagadeesh, B. (2009). Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli. *Cerebral Cortex*, *19*(4), 760–776.

Arsalidou, M., Morris, D., & Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. *Brain Topography*, *24*(2), 149–163.

Banich, M. T. (2003). The divided visual field technique in laterality and interhemispheric integration. In Kenneth Hugdahl (Ed.), *Experimental methods in neuropsychology* (pp. 47–63). Springer New York.

Bartlett, M. S. (2007). Information maximization in face processing. *Neurocomputing*, *70*(13–15), 2204–2217.

Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, *15*(2), 145–153.

Behrmann, M., & Plaut, D. C. (2015). A vision of graded hemispheric specialization. *Annals of the New York Academy of Sciences*, *1359*(1), 30–46.

Berisha, F., Johnston, A., & McOwan, P. W. (2010). Identifying regions that carry the best information about global facial configurations. *Journal of Vision*, *10*(11), 27–27.

Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience and Biobehavioral Reviews*, *55*, 536–546.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, *3*(8).

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

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99* (Vol. 2, pp. 187–194).

Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, *353*(6343), 429–431.

Bradshaw, J. L., & Nettleton, N. C. (1981). The nature of hemispheric specialization in man. *Behavioral and Brain Sciences*, *4*(1981), 52–91.

Bradshaw, J. L., Taylor, M. J., Patterson, K., & Nettleton, N. C. (1980). Upright and inverted faces, and housefronts, in the two visual fields: A right and a Left hemisphere contribution. *Journal of Clinical Neuropsychology*, *2*(4), 245–257.

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

Broman, M. (1978). Reaction-Time Differences Between the Left and Right Hemispheres for Face and Letter Discrimination in Children and adults. *Cortex*, *14*(4), 578–591.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327.

Bülthoff, H. H., Cunningham, D. W., & Wallraven, C. (2011). Dynamic Aspects of Face Processing in Humans. In *Handbook of Face Recognition* (pp. 575–596).

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "Spotlight" reveals universal information use for face recognition. *PLoS ONE*, *5*(3), 1–12.

Calder, A. J. (2011). Does facial identity and facial expression recognition involve separate visual routes? In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V Haxby (Eds.), *Oxford Handbook of Face Perception*. Oxford University Press.

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., & Henson, R. N. A. (2007). Separate Coding of Different Gaze Directions in the Superior Temporal Sulcus and Inferior Parietal Lobule. *Current Biology*, *17*(1), 20–25.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews. Neuroscience*, *6*(8), 641–651.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., … David, A. S. (1997). Activation of Auditory Cortex During Silent Lipreading. *Science*, *276*(5312), 593–596.

Calvert, G. A., & Campbell, R. (2003). Reading Speech from Still and Moving Faces: The Neural Substrates of Visible Speech. *Journal of Cognitive Neuroscience*, *15*(1), 57–70.

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lipreading: A neurological dissociation. *Brain*, *109*(3), 509–521.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*(4275), 312–314.

Carlin, J. D., Calder, A. J., Kriegeskorte, N., Nili, H., & Rowe, J. B. (2011). A head view-invariant representation of gaze direction in anterior superior temporal sulcus. *Current Biology*, *21*(21), 1817–1821.

Chan, A. W. Y. (2013). Functional organization and visual representations of human ventral lateral prefrontal cortex. *Frontiers in Psychology*, *4*(JUL), 1–15.

Chan, A. W. Y., & Downing, P. E. (2011). Faces and Eyes in Human Lateral Prefrontal Cortex. *Frontiers in Human Neuroscience*, *5*(June), 1–10.

Chang, L., & Tsao, D. Y. (2017). The Code for Facial Identity in the Primate Brain. *Cell*, *169*(6), 1013–1028.

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In Bad Taste: Evidence for the Oral Origins of Moral Disgust. *Science*, *323*(5918), 1222–1226.

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1327–1336.

Collins, J. A., & Olson, I. R. (2014). Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*, *61*(1), 65–79.

Cook, R., Aichelburg, C., & Johnston, A. (2015). Illusory Feature Slowing: Evidence for Perceptual Models of Global Facial Change. *Psychological Science*, *26*(4), 512–517.

Cortes, J. M., Marinazzo, D., Series, P., Oram, M. W., Sejnowski, T. J., & Van Rossum, M. C. W. (2012). The effect of neural adaptation on population coding accuracy. *Journal of Computational Neuroscience*, *32*(3), 387–402.

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, *111*(2), 219–247.

Cunningham, D. W., & Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *Journal of Vision*, *9*(13), 7–7.

Curio, C., Giese, M. A., Breidt, M., Kleiner, M., & Bülthoff, H. H. (2008). Probing dynamic human facial action recognition from the other side of the mean. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization - APGV '08* (p. 59).

Dakin, S. C., & Watt, R. J. (2009). Influence of case type, word frequency, and exposure duration on visual word recognition. *Journal of Vision*, *9*, 1–10.

de Gelder, B., Van den Stock, J., Meeren, H. K. M., Sinke, C. B. A., Kret, M. E., & Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd.

de la Rosa, S., Giese, M. A., Bülthoff, H. H., & Curio, C. (2013). The contribution of different cues of facial movement to the emotional facial expression adaptation aftereffect. *Journal of Vision*, *13(1)*(23), 1–15.

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere-An MRI and PET study and a review of the literature. *Neuropsychologia*, *32*(8), 893–902.

De Souza, W. C., Eifuku, S., Tamura, R., Nishijo, H., & Ono, T. (2005). Differential characteristics of face neuron responses within the anterior superior temporal sulcus of macaques. *Journal of Neurophysiology*, *94*(2), 1252–1266.

De Winter, F. L., Zhu, Q., Van den Stock, J., Nelissen, K., Peeters, R., de Gelder, B., … Vandenbulcke, M. (2015). Lateralization for dynamic facial expressions in human superior temporal sulcus. *NeuroImage*, *106*, 340–352.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology. General*, *115*(2), 107–117.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends*

in Cognitive Sciences, *11*(8), 333–341.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434.

Dien, J. (2008). Looking both ways through time: The Janus model of lateralized cognition. *Brain and Cognition*, *67*(3), 292–323.

Dien, J. (2009). A tale of two recognition systems: Implications of the fusiform face area and the visual word form area for lateralized object recognition models. *Neuropsychologia*, *47*(1), 1–16.

Draper, B. A., Baek, K., Bartlett, M. S., & Beveridge, J. R. (2003). Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding*, *91*(1–2), 115–137.

Duchaine, B., Dingle, K., Butterworth, E., & Nakayama, K. (2004). Normal greeble learning in a severe case of developmental prosopagnosia. *Neuron*, *43*(4), 469–473.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, *1*, 393–416.

Eifuku, S., De Souza, W. C., Tamura, R., Nishijo, H., & Ono, T. (2004). Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *Journal of Neurophysiology*, *91*(1), 358–371.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3), 169–200.

Ekman, P., & Friesen, W. V. (1976a). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75.

Ekman, P., & Friesen, W. V. (1976b). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists.

Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion*, *12*(3), 303–308.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-Cultural Elements in Facial Displays of Emotion. *Science*, *164*(3875), 86–88.

Ellis, H. D., & Shepherd, J. W. (1975). Recognition of Upright and Inverted Faces Presented in the Left and Right Visual Fields. *Cortex*, *11*(1), 3–7.

Engel, S. A. (2012). The development and use of phase-encoded functional MRI designs. *NeuroImage*, *62*(2), 1195–1200.

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, *45*(14), 3234–3241.

Ethofer, T., Gschwind, M., & Vuilleumier, P. (2011). Processing social aspects of human gaze: A combined fMRI-DTI study. *NeuroImage*, *55*(1), 411–419.

Fernández-Ruiz, J., & Díaz, R. (1999). Prism adaptation and aftereffect: Specifying the properties of a procedural memory system. *Learning & Memory*, *6*(1), 47–53.

Finlay, D. C., & French, J. (1978). Visual field differences in a facial recognition

task using signal detection theory. *Neuropsychologia*, *16*(1), 103–107.

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *Journal of Vision*, *11*(3), 1–15.

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637–1651.

Frässle, S., Paulus, F. M., Krach, S., Schweinberger, S. R., Stephan, K. E., & Jansen, A. (2016). Mechanisms of hemispheric lateralization: Asymmetric interhemispheric recruitment in the face perception network. *NeuroImage*, *124*, 977–988.

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.

Frisby, J. P., & Stone, J. V. (2010). *Seeing: The Computational Approach to Biological Vision* (2nd ed.). MIT Press.

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Friston, K. J., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221.

Frund, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, *11*(6), 16–16.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(June), 773–778.

Furl, N., Henson, R. N., Friston, K. J., & Calder, A. J. (2015). Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. *Cerebral Cortex*, *25*(9), 2876–2882.

Furl, N., Phillips, P. J., & O'Toole, A. J. (2002). Face recognition algorithms and the other-race effect: computational mechamisms for a developmental contact hypothesis. *Cognitive Science*, *26*, 797–815.

Furl, N., Van Rijsbergen, N. J., Kiebel, S. J., Friston, K. J., Treves, A., & Dolan, R. J. (2010). Modulation of perception and brain activity by predictable trajectories of facial expressions. *Cerebral Cortex*, *20*(3), 694–703.

Gauthier, I., Behrmann, M., & Tarr, M. J. (1999). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience*, *11*(4), 349–370.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition*, *103*(2), 322–330.

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, *6*(4), 428–432.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W.

(2000). The Fusiform "Face Area" is Part of a Network that Processes Faces at the Individual Level. *Journal of Cognitive Neuroscience*, *12*(3), 495–504.

Gauthier, 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*(6), 568–573.

Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? *Brain*, *123*, 1293–1326.

Geffen, G., Bradshaw, J. L., & Wallace, G. (1971). Interhemispheric effects on reaction time to verbal and nonverbal visual stimuli. *Journal of Experimental Psychology*, *87*(3), 415–422.

Giese, M. a, & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews. Neuroscience*, *4*(3), 179–192.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"--holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1023–1039.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, *6*(4), 316–328.

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, *45*(22), 2847–2853.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2012). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex*, *22*(7), 1564–1576.

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

Hagler, D. J., & Sereno, M. I. (2006). Spatial maps in frontal and prefrontal cortex. *NeuroImage*, *29*(2), 567–577.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45–56.

Haxby, J. V., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron*, *22*(1), 189–199.

Hayward, W. G., Crookes, K., & Rhodes, G. (2013). The other-race effect: Holistic coding differences and beyond. *Visual Cognition*, *21*(9–10), 1224–1247.

Hein, G., & Knight, R. T. (2008). Superior Temporal Sulcus—It's My Area: Or Is It?

Journal of Cognitive Neuroscience, *20*(12), 2125–2136.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*(11), 880–885.

Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. *Trends in Cognitive Sciences*, *13*(5), 216–221.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception*, *23*(1), 65–74.

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, *436*(7047), 71–77.

Huang, R., & Sereno, M. I. (2013). Bottom-up Retinotopic Organization Supports Top-down Mental Imagery. *The Open Neuroimaging Journal*, *7*(1), 58–67.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing Prejudice: Implicit Prejudice and the Perception of Facial Threat. *Psychological Science*, *14*(6), 640–643.

Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in Social Categorization. *Psychological Science*, *15*(5), 342–345.

Iwasaki, M., & Noguchi, Y. (2016). Hiding true emotions: micro-expressions in eyes retrospectively concealed by mouth movements. *Scientific Reports*, *6*(October 2015), 22049.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*(2), 187–192.

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*(19), 7241–7244.

Jack, R. E., & Schyns, P. G. (2015). The Human Face as a Dynamic Tool for Social Communication. *Current Biology*, *25*(14), R621–R634.

Johnston, A. (2011). Is dynamic face perception primary? In C. Curio, M. A. Giese, & H. H. Bulthoff (Eds.), *Dynamic faces: Insights from experiments and computation* (pp. 3–14). Cambridge, MA: MIT Press.

Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings. Biological Sciences / The Royal Society*, *250*(1329), 297–306.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*(1), B1–B11.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11).

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166.

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M. C., … Song, J. H. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*, *46*(3), 786–802.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(14), 1.

Krumhuber, E., Manstead, A. S. R., & Kappas, A. (2007). Temporal aspects of facial displays in person and expression perception: The effects of smile dynamics, head-tilt, and gender. *Journal of Nonverbal Behavior*, *31*(1), 39–56.

Kubota, J. T., & Ito, T. A. (2007). Multiple cues in social perception: The time course of processing race and facial expression. *Journal of Experimental Social Psychology*, *43*(5), 738–752.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology*, *12*(4), 259–272.

Laurence, S., Hole, G. J., & Hills, P. J. (2014). Lecturers' faces fatigue their students: Face identity aftereffects for dynamic and static faces. *Visual Cognition*, *22*(8), 1072–1083.

Leder, H., & Bruce, V. (1998). Local and Relational Aspects of Face Distinctiveness. *The Quarterly Journal of Experimental Psychology Section A*, *51*(3), 449–473.

Leder, H., & Carbon, C.-C. (2005). When context hinders! Learn-test compatibility in face recognition. *The Quarterly Journal of Experimental Psychology*, *58*(2), 235–250.

Leehey, S., Carey, S., Diamond, R., & Cahn, A. (1978). Upright and Inverted Faces: The Right Hemisphere Knows the Difference. *Cortex*, *14*(3), 411–419.

Lehky, S. R., & Sereno, A. B. (2007). Comparison of Shape Encoding in Primate Dorsal and Ventral Visual Pathways. *Journal of Neurophysiology*, *97*(1), 307–319.

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

Lewis, M. B. (2001). The lady's not for turning: Rotation of the Thatcher illusion. *Perception*, *30*(6), 769–774.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience*, *5*(9), 910–916.

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of Face Parts and Face Configurations: An fMRI Study. *Journal of Cognition*, *22*(1), 203–211.

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

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. WH San Francisco: Freeman and Company.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

McKone, E. (2010). Face and Object Recognition: How Do They Differ? In V. Coltheart (Ed.), *Tutorials in Visual Cognition* (pp. 261–303). Routledge.

McKone, E., & Robbins, R. A. (2007). The evidence rejects the expertise hypothesis: Reply to Gauthier & Bukach. *Cognition*, *103*(2), 331–336.

McKone, E., & Robbins, R. A. (2011). Are faces special? In G. Rhodes, A. J. Calder, M. Johnston, & J. V Haxby (Eds.), *The Oxford handbook of face perception* (pp. 149–176). Oxford University Press.

Meytlis, M., & Sirovich, L. (2007). On the Dimensionality of Face Space. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 29, pp. 1262–1267).

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, *335*(6193), 817–820.

Moeller, S., Freiwald, W. a, & Tsao, D. Y. (2008). Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science*, *320*(5881), 1355–1359.

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

Movshon, J. A., & Simoncelli, E. P. (2014). Representation of Naturalistic Image Structure in the Primate Visual Cortex. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 115–122.

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biology*, *3*(8).

Munafò, M. R., & Smith, G. D. (2018). Repeating experiments is not enough. *Nature*, *553*, 399–401.

Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Research*, *46*(15), 2456–2464.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, *18*(1), 55–68.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, *25*(2), 499–507.

Nestares, O., & Heeger, D. J. (2000). Robust multiresolution alignment of MRI brain volumes. *Magnetic Resonance in Medicine*, *43*(5), 705–715.

Nichols, D. F., Betts, L. R., & Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology*, *1*(JUL), 1–13.

Nishida, S., & Johnston, A. (2010). The time marker account of cross-channel temporal judgments. In R. Nijhawan & B. Khurana (Eds.), *Space and Time in Perception and Action* (pp. 278–300). Cambridge University Press.

O'Toole, A. J. (2011). Cognitive and Computational Approaches to Face Recognition. In A. J. Calder (Ed.), *Oxford Handbook of Face Perception*. Oxford University Press.

O'Toole, A. J., & Roark, D. A. (2011). Memory for moving faces: the interplay of two recognition systems. In C. Curio, B. H. H., & M. A. Giese (Eds.), *Dynamic Faces: Insights from Experiments and Computation* (pp. 15–29).

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*(6), 261–266.

O Scalaidhe, S. P. (1997). Areal Segregation of Face-Processing Neurons in Prefrontal Cortex. *Science*, *278*(5340), 1135–1138.

Ossowski, A., & Behrmann, M. (2015). Left hemisphere specialization for word reading potentially causes, rather than results from, a left lateralized bias for high spatial frequency visual information. *Cortex*, *72*(January), 27–39.

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

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London B: Biological Sciences*, *223*(1232), 293–317.

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature. *Frontiers in Psychology*, *3*(DEC), 1–11.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*(4), 2356–2363.

Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, *24*(17), 2066–2070.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, *209*(4), 481–493.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS Evidence for the Involvement of the Right Occipital Face Area in Early Face Processing. *Current Biology*, *17*(18), 1568–1573.

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847.

Purves, D., Monson, B. B., Sundararajan, J., & Wojtach, W. T. (2014). How biological vision succeeds in the physical world. *Proceedings of the National Academy of Sciences*, *111*(13), 4750–4755.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Reinl, M., & Bartels, A. (2014). Face processing regions are sensitive to distinct aspects of temporal sequence in facial dynamics. *NeuroImage*, *102*(P2), 407–415.

Reynolds, D. M., & Jeeves, M. A. (1978). A developmental study of hemispheric specialization for alphabetic stimuli. *Cortex*, *14*, 259–267.

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

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

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology*, *14*(23), 2119–2123.

Rhodes, G., & Leopold, D. a. (2011). Adaptive Norm-Based Coding of Face Identity. In A. J. Calder (Ed.), *The Oxford handbook of face perception* (pp. 263–286). Oxford University Press.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological Science*, *22*(4), 464–471.

Richler, J. J., & Gauthier, I. (2014). A Meta-Analysis and Review of Holistic Face Processing. *Psychological Bulletin*, *140*(5), 1281–1302.

Rizzolatti, G., Umilta, C., & Berlucchi, G. (1971). Opposite superiorities of the right and left cerebral hemispheres in discriminative reaction time to physiognomic and alphabetic material. *Brain*, *94*, 431–442.

Robbins, R. A., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79.

Ross, D. A., Deroche, M., & Palmeri, T. J. (2014). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin & Review*, *21*(1), 47–70.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *128*(2), 274–289.

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*(2), 139–253.

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature*

Reviews Neuroscience, *8*(7), 522–535.

Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of Visual Information Integration in the Brain for Categorizing Facial Expressions. *Current Biology*, *17*, 1580–1585.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, *14*(5), 391–396.

Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., … Tootell, R. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893.

Sergent, J. (1982). The Cerebral Balance of Power: Confrontation or Cooperation? *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 253–272.

Sergent, J., Ohta, S., & Macdonald, B. (1992). Functional Neuroanatomy of Face and Object Processing. *Brain*, *115*(September), 15–36.

Sigala, R., Serre, T., Poggio, T., & Giese, M. a. (2005). Learning features of intermediate complexity for the recognition of biological motion. *Artificial Neural Networks: Biological Inspirations–ICANN*, 241–246.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, *4*(3), 519.

Sporer, S. L. (2001). RECOGNIZING FACES OF OTHER ETHNIC GROUPS: An Integration of Theories. *Psychology, Public Policy, & Law March*, *7*(1), 36–97.

Sterzer, P., Haynes, J. D., & Rees, G. (2006). Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. *NeuroImage*, *32*(3), 1308–1316.

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

Stürzel, F., & Spillmann, L. (2000). Thatcher illusion: Dependence on angle of rotation. *Perception*, *29*(8), 937–942.

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

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225–245.

Tanaka, J. W., Giles, M., Kremen, S., & Simon, V. (1998). Mapping attractor fields in face space: the atypicality bias in face recognition. *Cognition*, *68*(3), 199–219.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, *25*(5), 583–592.

Tanaka, J. W., & Simonyi, D. (2016). The “parts and wholes” of face recognition: A review of the literature. *Quarterly Journal of Experimental Psychology*

(2006), *69*(10), 1876–1889.

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences*, *7*(1), 23–30.

Taubert, J., & Alais, D. (2009). The composite illusion requires composite face stimuli to be biologically plausible. *Vision Research*, *49*(14), 1877–1885.

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision Research*, *51*(11), 1273–1278.

Thompson, J. C. (2005). Configural Processing of Biological Motion in Human Superior Temporal Sulcus. *Journal of Neuroscience*, *25*(39), 9059–9066.

Thompson, P. (1980). Margaret Thatcher: A New Illusion. *Perception*, *9*(4), 483–484.

Tranel, D., Damasio, a R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology*, *38*(5), 690–696.

Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron*, *24*(1), 105–110.

Troje, N. F. (2013). Vision as hypothesis testing: The case of biological motion perception. In S. Leyssen & P. Rathgeber (Eds.), *Bilder animierter Bewegung / Images of Animate Movement*. Wilhelm Fink Verlag: Paderborn.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411–437.

Tsao, D. Y., Moeller, S., & Freiwald, W. a. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19514–19519.

Tsao, D. Y., Schweers, N., Moeller, S., & Freiwald, W. a. (2008). Patches of face-selective cortex in the macaque frontal lobe. *Nature Neuroscience*, *11*(8), 877–879.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373.

Turk, M., & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.

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

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space : A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology*, *69*(10), 1996–2019.

Wada, Y., & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region.

Journal of Neurology, Neurosurgery, and Psychiatry, *71*(2), 254–257.

Wallis, G., & Baddeley, R. (1997). Optimal, unsupervised learning in invariant object recognition. *Neural Computation*, *9*, 883–894.

Wallis, G., & Bülthoff, H. H. (1999). Learning to Recognize Objects. *Trends in Cognitive Sciences*, *3*(1), 22–31.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4800–4804.

Webster, M. A. (2015). Visual adaptation. *Annual Review of Vision Science*, *1*(1), 547–569.

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

Weiner, K. S., & Grill-spector, K. (2012). The improbable simplicity of the fusiform face area. *Trends in Cognitive Sciences*, *16*(5), 251–254.

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97.

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

Xu, H., Liu, P., Dayan, P., & Qian, N. (2012). Multi-level visual adaptation: Dissociating curvature and facial-expression aftereffects produced by the same adapting stimuli. *Vision Research*, *72*(December), 42–53.

Yassa, M. A., & Stark, C. E. L. (2009). A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *NeuroImage*, *44*(2), 319–327.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145.

Young, A. W. (2011). Disorders of face perception. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V Haxby (Eds.), *Oxford Handbook of Face Perception*. Oxford University Press.

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

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research*, *48*(2), 63–68.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, *15*(24), 2256–2262.

Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. In *Computers and Graphics (Pergamon)* (Vol. 36, pp. 152–162). Elsevier.

Zhang, J., Li, X., Song, Y., & Liu, J. (2012). The fusiform face area is engaged in holistic, not parts-based, representation of faces. *PLoS ONE*, *7*(7).

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research*, *105*, 61–69.