This chapter appears as Chapter 13 in 'What is adaptive about adaptive memory’, to cite:

INTRODUCTION

Humans are social and we have evolved in close proximity with one another, surrounded by different identities almost everywhere we turn. The accurate perception of others is a fundamental aspect of social cognition, allowing us to detect the intention, attention and identity of an individual (among other attributes) (Bruce & Young, 1986). Indeed humans have developed an expertise for face perception, which arguably exceeds the perception of other stimuli: we can remember thousands of faces as 'unique'. When we consider the influence that this skill has on our behaviour it is not difficult to understand its evolutionary significance and how it is likely to have assisted in negotiating our survival, mediating among other things, approach-avoidance behaviour (Engell, Haxby, & Todorov, 2007; Todorov, 2008), mate selection (Little, Jones, & DeBruine, 2011; Rhodes, 2006; Rhodes, Simmons, & Peters, 2005), the recognition of familiar, unrelated others (Bruce, Henderson, Newman, & Burton, 2001; Burton, Wilson, Cowen, & Bruce, 1999; Rossion, Schiltz, & Crommelinck, 2003) and the recognition of kin (DeBruine et al., 2009; Maloney & Dal Martello, 2006). Yet for such a skill to be truly adaptive we must consider that recognising others is not purely dependent on our memory for faces alone but that other social information, particularly information in the voice, can be used to identify others. Like the face, the voice conveys information which can act as an identity signature and affect recognition and thus can be thought of as an ‘auditory face’ (Belin, Fecteau, & Bédard, 2004).

In this chapter we explore the selection pressures that were likely to have mediated the development of our ability to recognise others. We consider how face recognition reflects a unique cognitive process with genetic underpinnings, highlighting face recognition as a phylogenetic adaptation (i.e. within species adaptation). Although recent research has highlighted the influence of genes on face recognition abilities, this chapter discusses evidence that face recognition abilities are rapidly acquired, and can be influenced by our surrounding environment. This experience-dependent malleability suggests that face recognition is also a product of ontogenetic adaptation (i.e. individual development across the lifespan), which has direct consequences on our ability to discriminate and remember faces. Although most research on face perception has been based on the recognition of faces from static images, here we also discuss research which highlights our ability to recognise faces under naturalistic conditions, such as the face in motion. Finally, we review the neural processes underlying face recognition and how they may share similar neural underpinnings to voice recognition, with the argument that such common processing of face and voice information most likely evolved to support person recognition throughout the lifespan.

FACE RECOGNITION: IT'S IN OUR GENES

Genetic fitness: recognising faces

Faces are arguably the most dominant visual stimuli in our surroundings and we, in turn, spend a vast amount of our time perceiving faces in order to gather relevant social information. Even when we are not specifically attending to faces they capture our attention (Langton, Law, Burton, & Schweinberger, 2008). Our fascination with faces is not surprising when
we consider the array of social signals that the face conveys, including the identity, intent and attractiveness of an individual. Our perceptual system has evolved with the capacity to efficiently extract this information to assist in interacting with our social world (Leopold & Rhodes, 2010). Indeed, the emergence of such a skill is likely to have facilitated our survival and to have developed in line with one of the most dominant selection pressures, that of sexual selection. Our ability to effectively deduce differences across individual faces has assisted in determining the genetic fitness of mate choices. Indeed it is argued that across cultures we have evolved a common percept of what facial characteristics are deemed attractive (Little et al., 2011). These biological standards of beauty include sexual dimorphism (femininity cues in female faces or masculinity cues in male faces), bilateral symmetry (how similar each half of the face is to the other) and averageness (facial features which are determined as the mathematical average for a given population) (for reviews see Little et al., 2011; Rhodes, 2006) and are suggested to be expressions of genetic fitness cues including fertility (Johnston, 2000; Johnston, Hagel, Franklin, Fink, & Grammer, 2001; Law Smith et al., 2006) and immunocompetence (Lie, Rhodes, & Simmons, 2008). See Figure 1 for illustration of how averageness can enhance perceived attractiveness.

For example, it has been suggested that visual features indicating femininity in a female face (short, narrow jaw and fuller lips) serve as a cue to fertility throughout the lifespan (Johnston, 2000; Johnston & Franklin, 1993; Law Smith et al., 2006; Perrett et al., 1998). In addition, we rate as both healthier and more attractive the faces of individuals associated with ‘good quality’ immune genes, compared to those who are not (Roberts, Little, Gosling, Perrett, et al., 2005). These genes, which are thought to be important in mate selection, are called major histocompatibility complex (MHC) genes and are associated with immune defence. However, in order to avoid inbreeding and to enhance the immunocompetency of offspring it has been suggested that we should prefer individuals who exhibit a MHC combination which is different from our own. For example, often mating patterns are observed to be non-random and reflect either ‘assortative’-similar genotypes/phenotypes in mating partners or ‘disassortative’ choices-dissimilar genotypes/phenotypes in mating partners and can be understood under the genetic compatibility hypothesis (Neff & Pitcher, 2005). However, faces of individuals who possess dissimilar immune genes are not necessarily those that are rated as being highly attractively (Roberts, Little, Gosling, Jones, et al., 2005; Roberts, Little, Gosling, Perrett, et al., 2005), suggesting that disassortative mating is not the only influence on preferences for faces.

However, other perceptual studies demonstrate that we do use facial cues of kinship (familial resemblance) to effectively avoid inbreeding which may have a negative effect on genetic fitness. For example, when our own face is morphed with an opposite sex face we tend to find the resulting face less attractive than if it is morphed with a same sex face. However participants who share no genetic relatedness to the face images do not show this pattern for the same and opposite sex face images (DeBruine, Jones, Little, & Perrett, 2008; DeBruine, 2004a). Moreover, this is particularly true for opposite sex siblings,
for example women who have brothers find the faces of ‘self-resembling’ unfamiliar male faces less attractive than women who do not have brothers (DeBruine et al., 2011). However, if there is no identity change in face images we find the faces of male and female ‘siblings’ similarly attractive, (for example, see Figure 4 in Bülthoff & Newell, 2004), suggesting that features that are attractive are not sex-specific (Casey, Mernagh, & Newell, 2009).

Faces that are deemed attractive induce activity in areas of the brain associated with reward (O’Doherty et al., 2003). Attractive faces have also been suggested to be better remembered than their less attractive counterparts (Marzi & Viggiano, 2010; Tsukiura & Cabeza, 2011), however this finding is not always found (Light, Hollander, & Kayra-Stuart, 1981; Wickham & Morris, 2003). Thus, face recognition serves a largely adaptive function, as we can accurately deduce cues from the face that may assist in increasing genetic transmission by producing healthy offspring. Evidence suggests that the perceived attractiveness of a face shares a relationship with sexual behaviour, increasing mating success (Rhodes et al., 2005). Although these traits are preferred across cultures, it is important to note that they vary in line with our familiarity to faces. For example, average faces are deemed more attractive when they reflect the average of our own group prototype rather than another group (Potter & Corneille, 2008) and increasing exposure to a face also increases its perceived attractiveness (Peskin & Newell, 2004).

As discussed, face recognition appears to affect mate selection and in turn face recognition also endows us with the ability to identify our own kin and we are adept at determining the genetic relatedness of individual faces (DeBruine et al., 2009; Maloney & Dal Martello, 2006). DeBruine (2004b) reported that we tend to find the faces of infants who resemble ourselves as more appealing than those who do not, even when we are unaware that a similarity may exist. Here, DeBruine (2004b) transformed participants faces to reflect neotenous face proportions and then morphed this face with that of an unfamiliar infant morph, such that the resulting infant face morph did not appear older through the inclusion of self-related facial information. There is evidence to suggest that this effect may be more evident in male than female participants as men may rely heavily on facial likeness as a verification of paternity (Platek, Burch, Panyavin, Wasserman, & Gallup, 2002). Other research suggests that the faces of newborn infants also draw attention (Brosch, Sander, & Scherer, 2007). Together these findings propose a largely adaptive function, when we consider that protecting infants is fundamental to reproductive success (Brosch et al., 2007).

**Genetic underpinnings: recognising faces**

The abilities discussed above, including kin recognition, are largely harnessed by our capacity to detect often subtle differences between faces and this is particularly true when determining the unique identity of an individual face. Although the process appears effortless, recognising and remembering faces as individual is a challenge for the perceptual system. Each face we encounter shares the same basic visual-feature organisation as the next, yet we can accurately discriminate between these faces. Furthermore we can
recognise the face of a familiar person even over large changes in visual image appearance, such as viewpoint, expression or illumination (Bruce, 1982; Bruce et al., 2001; Burton et al., 1999) (see Figure 2). Recent findings suggest that our capacity for remembering faces is robustly separable from other general cognitive abilities, including measures of immediate and delayed recall (Wilhelm, Herzmann, Kunina, & Sommer, 2007). In addition, face recognition performance appears to follow a unique cognitive trajectory, continuing to improve until just beyond the age of 30 years. In contrast other cognitive skills associated with person recognition, for example name recognition, cease to show age improvements almost a decade earlier (Germine, Duchaine, & Nakayama, 2011).

Recent research has suggested that the ability to recognise other faces varies across individuals. This variation is inherently dissociable from individual differences in other general cognitive abilities (Herzmann, Kunina, Sommer, & Wilhelm, 2010; Wilhelm et al., 2010). Differences in face recognition performance are so evident across individuals that they have been shown to be quantitatively measureable (Russell, Chatterjee, & Nakayama, 2012). For example, recent evidence has suggested the existence of ‘super-recognizers’: individuals with extraordinary face recognition ability which greatly surpasses average face recognition performance (Russell, Duchaine, & Nakayama, 2009). In fact, the degree of enhanced performance observed in these individuals is quantitatively similar to the degree of deficit expressed in face recognition performance in developmental prosopagnosics (also referred to as congenital prosopagnosia) (Russell et al., 2009). Developmental prosopagnosics are individuals who have a specific impairment in face recognition performance that is not attributable to brain trauma or insult (see further discussion below).

Several studies have suggested that rather like the highly heritable and neurobiological correlates that have been observed for general intelligence (for review see Toga & Thompson, 2005), the degree of expertise expressed in face perception may also be largely genetically determined. This genetic component suggests that as humans, face recognition ability may have largely evolved as a phylogenetic adaptation (within species adaptation) (Johnson, 2011). The most compelling evidence that genetics feature as a contributor in face recognition ability come from the results observed through twin studies (Wilmer et al., 2010; Zhu et al., 2010). What is important about these studies in the current context is that the role of genetics can studied while also controlling for environmental influences, as twins who are raised together largely share similar environmental factors, in the same way as non-identical twins or other siblings. Wilmer et al. (2010) observed that monozygotic (MZ) twins, (i.e. developed from one fertilized oocyte) show highly correlated face recognition performance. In fact the correlation between their performance was more than double the correlation observed for dizygotic (DZ) twins, (i.e. developed from different fertilised oocytes), raised in the same environment (correlation 0.70 versus 0.29, respectively). As the MZ and DZ twin sets both shared similar environmental factors, the observed differences can be assumed to largely reflect genetic underpinnings (Wilmer et al., 2010). However, despite
these recent findings that the ability to perceive faces is influenced by genetics, the associated adaptive function of these individual differences is, for the moment, unclear.

Zhu et al. (2010) provided similar evidence for face recognition ability being highly correlated for MZ and less so for DZ twins raised together. Intriguingly, they found no evidence for such a relationship for object recognition abilities across twin sets. They did, however, demonstrate a similar pattern of correlational strength for the ‘face inversion effect’ across twin sets. Image inversion has been shown to disrupt face recognition, yet it is thought to elicit less disruption in the recognition of other objects and is suggested to reflect a method of processing, holistic or configural processing, which is largely specific to face perception (Yin, 1969; Young, Hellawell, & Hay, 1987; Maurer, Le Grand, & Mondloch, 2002). Thus the effect Zhu et al. (2010) observed for face recognition performance suggests that it is reflective of the perceptual processes that are recruited specifically for face recognition and not generic low level processing differences across individuals.

However, it is also important to note that not all of the variance in face recognition performance in the aforementioned twin studies was explained by genetic similarity. The statistical models used in such twin studies (Rabe-Hesketh, Skrondal, & Gjessing, 2008) take into consideration an error measurement which includes the contribution of non-familial influences i.e. the environmental influence unique to each twin. Thus although the ability to recognise faces may reflect a unique and highly heritable perceptual ability, environmental influences were also reported to be an important contributor in face recognition performance (a more in-depth discussion of the role of environmental factors on face recognition ability follows later in the chapter). What perhaps is intriguing however is that the effect of genetics on face recognition performance was more likely to explain the variance in face recognition performance for the older age group of twins (13-19 years) than for the younger age group (7-12 years) (Zhu et al., 2010). Thus although environmental factors have a role to play in face recognition ability, the older we become then the more likely it is that variability in facial recognition performance across individuals may relate to familial influence. For example, Burton et al. (1999) demonstrated that even police officers who were practiced in forensic identification (average of 13.5 years’ service) showed similar performance in recognising unfamiliar faces as untrained control participants, suggesting that individual differences in face recognition performance may not reflect ‘learned’ abilities. As such, the greater the genetic similarity between twins, as is the case for MZ twins, the more likely they will share a similar ability for recognising other individuals’ faces.

In addition to twin studies, our understanding of the genetic basis of face recognition has also been informed by the study of developmental prosopagnosia. Prosopagnosia (‘face blindness’) refers to a condition first documented by Bodamer in 1947 (see Ellis & Florence, 1990 for translated paper) in which patients demonstrated a distinct deficit in recognising individual faces, whereas other visual abilities remained intact. Beyn and Knyazeva (1962) described this condition as
reflecting a "failure in recognition at its most highly specific and individual level" (p158). While the first cases documented largely reflected acquired prosopagnosia, which results from insult to, and dysfunction of, the cortical face processing network (Barton, 2008; Busigny et al., 2010; Rossion et al., 2003; Schiltz et al., 2006), later studies demonstrated the existence of developmental prosopagnosia. Developmental prosopagnosics have atypical face processing, including impairment in recognising the identity of individual faces, and such deficits emerge in the absence of insult or atypical neurodevelopment (Lee, Duchaine, Wilson, & Nakayama, 2010). Developmental prosopagnosia has been shown to be largely heritable, with multiple cases observed in the same family (Duchaine, Germine, & Nakayama, 2007; Grueter et al., 2007; Lee et al., 2010; Schmalzl, Palermo, & Coltheart, 2008). Such findings may implicate the existence of an evolved neural substrate, largely located in the ventral stream of the cortex, for the recognition of faces (Young et al., 2008) with genetics arguably playing a role in both typical and atypical development.

FACE RECOGNITION: DEVELOPMENTAL AND ENVIRONMENTAL INFLUENCE

That face recognition evolved as a phylogenetic adaptation is also evident from the infant literature which suggests that newborns exhibit a preference for looking at ‘face-like’ stimuli (Fantz, 1963; Valenza, Simion, Cassia, & Umiltà, 1996) and for visually tracking stimuli which schematically represent a face, versus those that do not (Goren, Sarty, & Wu, 1975; Johnson, Dziurawiec, Ellis, & Morton, 1991). Such evidence suggests that the newborn perceptual system may be biased to attend to and process ‘face-like’ stimuli (Johnson, 2005) (see Figure 3 for example of stimuli). There is also evidence to suggest that even with such an immature visual system infants show a preference for their mother’s face over that of a stranger (Pascalis, de Schonen, Morton, Deruelle, & Fabre-Grenet, 1995), however this may be mediated by single-feature processing, such as hairstyle (Pascalis et al., 1995).

Although such precocious face perception abilities may reflect phylogenetic adaptation, it is also important to consider that faces are arguably the most dominant stimulus in the infant’s environment (Sinha, Balas, & Ostrovsky, 2007). As such, with the multitude of faces in the infant’s surrounding world, face information may be rapidly acquired, in turn, biasing the perceptual preferences of infants. Indeed for such a skill to be truly adaptive it must have evolved to accommodate and become tuned to the specific pressures of the surrounding environment. Evidence also suggests that face recognition is malleable to change, showing adaptive changes in line with ontogenetic individual developmental experience (Johnson, 2011). These adaptive changes are evident across the lifespan, being largely mediated by our degree of exposure to different face types. Perhaps two of the best examples of such adaptation are highlighted by the other-race effect (for a review, see Meissner & Brigham, 2001) and the own-age bias (for a review, see Rhodes & Anastasi, 2011). These influences are discussed in detail in the following two sections.
As noted, we are experts at recognising and remembering faces, and we can remember thousands of faces as unique. Yet, as impressive as this ability is, our memory for faces is also susceptible to a significant degree of error when the face we are to recognise belongs to an ethnic or other-race group with which we are unfamiliar. We perceive individual faces of other-race faces as sharing a large degree of similarity (Vizioli, Rousselet, & Caldara, 2010) and as a consequence exhibit increased ‘false alarm’ rates in memory for other-race faces in comparison with faces of the race by which we are familiar (Meissner & Brigham, 2001). This phenomenon is known in the literature as the ‘other-race effect’ or ‘own-race bias’ (Meissner & Brigham, 2001) and it emerges during the early stages of development through the process of perceptual narrowing (Kelly et al., 2007). For example, 3 month old infants can effectively discriminate between faces of both their own-race and other-race faces. However, by the age of 9 months the infant perceptual system has ‘narrowed’ and the ability to efficiently discriminate between two faces is restricted to own-race faces (Kelly et al., 2007). This own-race bias is mediated by exposure to same race faces in the infant’s environment. This exposure has direct consequences on our understanding of how we represent faces in memory (Valentine, 1991), by forming representations based on a prototype which is most beneficial for interacting within our surrounding social environment. Chiroro, Tredoux, Radaelli and Meissner (2008) demonstrated how finely tuned this ability is. They observed that white and black South Africans demonstrate an own-race bias only for the faces of the ethnic groups with which they have perceptual experience i.e. South African faces. When presented with faces of white and black Americans, they failed to observe enhanced recognition for ‘own-race’ faces (Chiroro et al., 2008). What this suggests is that the perceptual system is highly sensitive to the degree of variation even within what may be classed as ‘same-race’ faces. This mechanism is likely to be largely adaptive when we consider that our perceptual system becomes tuned to detecting and remembering the faces of those we are surrounded by and in turn more likely to encounter in the future. As such the false alarm rates for other-race faces can be equated to false memories observed in other memory tasks emphasizing that due to ontogenetic adaptation, our memory for recognising faces has most likely evolved to use past information to effectively predict future encounters (Nairne & Pandeirada, 2008).

As noted, face recognition performance appears to peak at just beyond the age of 30 years (Germain et al., 2011). Subsequently, performance appears to decline as a function of increasing age, with decline evident at around 50 years of age (Bowles et al., 2009). This decrease in face recognition performance appears to be specific to the recognition of faces in comparison to the recognition of other within-category objects (Boutet & Faubert, 2006). What is important to note is that although such changes are evident, they are dissociable from changes in general cognitive ability that occur with the ageing process. Throughout the lifespan, even until old age, face recognition appears to
represent a unique cognitive ability (Hildebrandt, Wilhelm, Schmiedek, Herzmann, & Sommer, 2011). These deficits in face recognition performance have been suggested to be mediated by a number of factors including poorer sensory acuity typically associated with the ageing process (Barnes, De L’Aune, & Schuchard, 2011; Norton, McBain, & Chen, 2009; Owsley, Sekuler, & Boldt, 1981), changes in how faces are encoded (Chaby, Narme, & George, 2011; Firestone, Turk-Browne, & Ryan, 2007; Gao et al., 2009; Grady et al., 1995; Murray, Halberstadt, & Ruffman, 2010), and marked differences in the underlying neural activity associated with processing faces (Grady, 2002; Haxby et al., 1994; Lee, Grady, Habak, Wilson, & Moscovitch, 2011), including less lateralised activation in face processing areas (Lee et al., 2011).

Perhaps one of the most striking behavioural effects which demonstrate this age related decline, is the higher level of ‘false alarm’ memory responses that older adults report during face recognition tasks, compared to younger adults (Bartlett & Leslie, 1986; Bartlett, Leslie, Tubbs, & Fulton, 1989; Rhodes, Castel, & Jacoby, 2008). For example, older adults show a bias for stating that they find a novel face familiar even when they have not encountered the face before. Although such findings may be explained by the underlying changes in sensory acuity and neuronal activity mentioned previously, it is also important to consider that older adults have been exposed to a larger number of faces across the lifespan than younger adults. This larger ‘database’ of faces in memory (or what Valentine (1991) referred to as ‘face-space’) may elicit more false alarm responses. By increasing the number of typical faces in face memory, newly encountered faces may be likely to elicit stronger signals of resemblance to another face within this large face-space, thus leading to false recognition. This may be one key reason why, as we get older, our ability to uniquely identify a face may appear to decline.

However, most studies that have reported a decline in face recognition performance in older compared to younger adults, have largely used images of younger adult faces as stimuli. Studies that have used face images of older adults as stimuli have reported that age related decline in face recognition is less evident for the recognition of older adult faces, relative to younger adult faces (Firestone, Turk-Browne & Ryan 2007; Fulton & Bartlett, 1991; Lamont, Stewart-Williams, & Podd, 2005). This effect is reported as the ‘own-age bias’ (for review, see Rhodes & Anastasi, 2011) and is similar to the other-race effect in that it is thought to emerge due to the extensive exposure that an individual has with faces of their own age group. This suggests that throughout the lifespan, even until old age, face recognition is malleable to effects such as the degree of exposure to particular faces in our surrounding environment.

Facing pressure: how social information affects face recognition

As illustrated by the own-age bias and other-race effect, face recognition appears to be tailored to facilitate optimal interaction with our social environment. Recent evidence has suggested that ‘high-level’ social information can in turn affect the ‘low-level’ perceptual processing of faces. For example, Bliss-Moreau, Barrett and Wright (2008) demonstrated that
information which assists in determining whether an individual may be helpful or threatening can directly affect how a face is perceived. They observed that unfamiliar faces, displaying a neutral expression, were more likely to be perceived as ‘negative’ or ‘positive’ when they had been respectively paired with statements with negative or positive valence, versus neutral statements. This effect was particularly evident when the statements were social in nature i.e. the behaviour directly referenced another individual. Furthermore, faces paired with such negative social information have been shown to dominate visual awareness, over faces paired with neutral or positive statements (Anderson, Siegel, Bliss-Moreau, & Barrett, 2011). In addition, it appears that such effects may also be evident for our memory for faces. For example, Chiappe et al. (2004) and Mealey, Daood, & Krage, (1996) observed that the faces of ‘cheaters’ (faces accompanied with cheating related information at the learning stage) were subsequently better remembered than faces paired with ‘trustworthy’/’co-operative’ or neutral information. However, others have suggested that the faces of ‘cheaters’ and ‘co-operative’ individuals may be similarly remembered, particularly when we are primed to process the presence of such behaviours (Felisberti & Pavey, 2010). Such findings suggest that without having direct experience with individual people, we use information gathered by others to assist in determining who may be helpful or threatening (Anderson et al., 2011) and that this in turn directly impacts our memory and perception for faces.

It may be argued that face perception when influenced by social information may reflect processes related to survival and as such our memory system may be biased towards remembering fitness relevant information (Nairne & Pandeirada, 2008; Nairne, Thompson, & Pandeirada, 2007; Nairne & Pandeirada, 2008). Indeed it has been documented that survival processing i.e. processing material, such as words, within the framework of a survival scenario, in comparison to a control scenario enhances recall (Nairne et al., 2007). However, research findings using the classic survival processing paradigm have failed to demonstrate similar survival processing enhancement on our memory for faces (Savine, Scullin, & Roediger, 2011). For example, when Savine et al. (2011) had participants rate faces for perceived helpfulness under primed ‘survival’ (e.g. lost in a desert and in danger of attack from predators) versus non-survival scenarios, they failed to observe a relative memory enhancement for faces viewed in the ‘survival’ scenarios. Thus, for the moment it remains unclear whether our memory for faces may be modulated by general survival information (Savine et al., 2011). However as discussed, many reports suggest that more specific survival relevant social information, can directly impact upon our perception of and memory for faces (Anderson et al., 2011). Intriguingly, recent evidence suggests that fitness related information in faces which may affect mate choice (sexual dimorphism in male faces), enhances recall of additionally presented non-face information, such as objects (Allan, Jones, DeBruine, & Smith, 2012).

FACE RECOGNITION: UNDERLYING NEURAL PROCESSES

Many researchers have argued for domain specificity in face processing, and that the cognitive processes involved in face
recognition appear to be distinct and specific for processing faces in comparison with other stimuli (Duchaine, Dingle, Butterworth, & Nakayama, 2004; McKone, Kanwisher, & Duchaine, 2007). This remains controversial, however, because others have suggested that face processing reflects a particular type of perceptual expertise in discriminating between similar visual stimuli, as discussed below. Evidence for domain specificity largely originated from neuropsychological investigations that reported selective deficits to face processing following damage to particular regions of the temporal lobe of the brain, which are involved in the processing of faces. Subsequent functional magnetic resonance imaging (fMRI) revealed a face module, known as the fusiform face area (FFA), which is located in visual areas of the ventral stream and is selectively activated when viewing images of faces (Kanwisher, McDermott, & Chun, 1997). In particular, this area is activated when viewing face images over images of other non-face objects or scrambled face images (suggesting that the area is not simply responsive to visual qualities of the image e.g. luminance). This area is not only active when we are perceiving faces but also when we are imagining them (O’Craven & Kanwisher, 2000) and electrical stimulation of this region specifically distorts the perception of faces over other stimuli (Parvizi et al., 2012). Moreover, the FFA is thought to be part of a larger neural network involved in the processing of faces, including the occipital face area (OFA) and areas in the superior temporal sulcus (STS) (Haxby, Hoffman, & Gobbini, 2000).

Although some have argued that this face sensitive activation may reflect ‘domain specificity’ (Kanwisher et al., 1997) for the processing of faces, it is important to note that the FFA has been suggested to also reflect an area recruited for within-category visual discrimination of other non-face stimuli, which require a similar level of visual ‘expertise’ for individual discrimination (Gauthier, Skudlarski, Gore, & Anderson, 2000). This ‘expertise hypothesis’ (Gauthier & Nelson, 2001; Gauthier et al., 2000) suggests that the observed cortical specialization for faces is a product of perceptual exposure and training in discriminating between similar visual stimuli. For example, it has been shown that individuals who are experts (or who are trained to become experts) in within-category discrimination of non-face stimuli, such as cars or birds, show similar underlying neural activity when perceiving faces or other ‘expertise’ stimuli (Gauthier et al., 2000; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002). In addition, highly familiar non-face stimuli that require holistic processing due to their multipart visual organisation (Bilalic, Langner, Ulrich, & Grodd, 2011) elicit activation in face selective areas. However, this finding remains controversial and a number of studies have reported greater neural activation to faces in comparison with objects of expertise in the FFA (Grill-Spector, Knouf, & Kanwisher, 2004; McKone et al., 2007). It is also important to note that expertise in face recognition and the involvement of the FFA in face processing appear to develop implicitly across all neurotypical individuals. On the other hand, the involvement of the FFA in discriminating between other ‘similar’ stimuli often involves explicit perceptual discrimination training. Moreover, others have argued that normal face recognition does not rely on the development of normal object recognition, as might be assumed from the
expertise hypothesis (Germine, Cashdollar, Düzel, & Duchaine, 2011).

However a bridging of these two viewpoints may be realised through the ‘interactive specialization’ approach (Johnson, 2011). Interactive specialization considers both the predisposition for cortical activation to e.g. face stimuli (intrinsic factors) and the level of exposure to faces (extrinsic factors), which modulates the degree of localization and specialization of this cortical activity. Thus, under this model areas of the brain may start out as having broadly tuned functions and with development they gradually become more refined for more specialized functions (Johnson, 2000, 2001), with direct consequences on localization and specialisation of neural activation in the brain. Several developmental studies have reported that children show differential neural activation to face stimuli in comparison to adults. For example, activation in face specific areas of the brain of children when viewing face stimuli is typically not observed, as the areas are relatively immature in terms of both specialisation of function and volume (Scherf, Behrmann, Humphreys, & Luna, 2007). When activation is observed in face selective areas, it is often much less localised than it is in the adult brain (Passarotti et al., 2003). Passarotti et al. (2003) stated that the neural systems involved in face processing continue to experience changes in terms of their functional specialization into late childhood. This account of increasing specialization in brain function is also in line with behavioural evidence supporting the idea that face recognition ability continues to improve into adulthood (Germine et al., 2011).

BEYOND THE FACE: PERSON RECOGNITION IS MULTISENSORY

As demonstrated throughout this chapter much research has been dedicated to understanding both the perceptual and neurological processes which underlie human face recognition. Although this has brought us far in understanding how the brain processes facial identity, the contribution of such studies to our understanding of person recognition in the real world is limited to information in the visual domain and more specifically, to the use of static images of faces as stimuli. Such investigations belie the true nature of the processes involved in the recognition of individuals we encounter every day, as faces move and speak, and are thus inherently dynamic, multisensory stimuli.

Dynamic faces

In the real world we rarely experience a face in static form and most often see the face in motion, across a large range of viewpoints and expression changes. It would seem odd if we encountered someone who did not move their face. Indeed, static face images extracted from moving sequences are viewed as less appealing than their dynamic counterparts, a finding referred to as the ‘frozen face’ effect (Post, Haberman, Iwaki, & Whitney, 2012). In addition, it has been reported that most regions of interest (ROI) involved in the core and extended face processing system are more successfully located through the use of dynamic (in comparison to static) face images in functional localizers (Fox, Iaria, & Barton, 2009).

The bias in using static images in studies of face perception has been
suggested to reflect an underlying assumption about face perception itself: that recognition is based on our ability to recognise the face through invariant characteristics whilst ignoring changes which occur through movement (Bruce & Young, 2012). Yet evidence from an emerging literature suggests that our ability to recognise a face may depend on our ability to process these dynamic changes and in turn create a more robust representation of the face in memory. For example, the way someone moves their forehead when in thought, may be peculiar to that individual and thus we may recognise the person not in spite of this movement, but perhaps in light of it. Thus, spatio-temporal information associated with an individual’s face may act as a unique signature to facilitate the recognition of that person (Bülthoff, Cunningham, & Wallraven, 2011; Knappmeyer, Thornton, & Bülthoff, 2003).

Although the face can move in numerous ways, movements can be largely classed as either rigid or non-rigid motion (O’Toole, Roark, & Abdi, 2002). Rigid motion refers to whole head movements e.g. rotation, nodding, whereas non-rigid motion refers to movements made through deformations of the face, i.e. the internal feature movement observed when making expressive movements or when speaking (Knappmeyer et al., 2003; O’Toole et al., 2002). Although the role for both types of information in learning new faces remains somewhat unclear (Christie & Bruce, 1998; Lander & Bruce, 2003; Lander & Chuang, 2005; Pike, Kemp, Nicola, & Phillips, 1997), what is apparent is the beneficial role that non-rigid motion plays in the recognition of familiar others, particularly under degraded visual conditions. Indeed recovering identity information from motion may be necessary for recognition under challenging viewing environments when facial form information may be compromised e.g. in dim lighting or at a distance. Under these conditions motion may help in determining if a face is familiar or unfamiliar. For example, Knight and Johnston (1997) demonstrated that when visual images of famous faces were degraded by presenting the images in their negative photograph form, participants were more accurate at identifying the target face when it was shown as a video clip rather than a static image. Furthermore they observed this dynamic enhancement to be larger for upright in comparison with inverted faces. Therefore the enhancement is not solely due to low level motion signals but relates to a ‘motion encoding system’, which is refined to upright faces (Hill & Johnston, 2001). Lander, Christie and Bruce (1999) similarly demonstrated that dynamic information can assist in the recognition of famous faces seen under suboptimal viewing conditions, including negatives, ‘threshold’ (one-bit per pixel black and white images) or inverted images, although the facilitation effect on inverted faces was small in comparison with other viewing conditions.

What is also evident is that preserving characteristic motion cues affects performance when learning new faces (Wallis, 1998; Wallis & Bülthoff, 2001) and recognising familiar others (Lander et al., 1999). That is, motion presented in forward form is more beneficial for the perceptual system than disrupted motion (e.g. reversed, or scrambled). This can be understood in terms of the temporal-association hypothesis (Wallis & Bülthoff, 1999). Under this hypothesis the visual system learns to associate multiple views of
for example a face when they are not only physically similar but also *temporally correlated* (Wallis & Bülthoff, 2001). Thus, if multiple views of a face are seen in close temporal proximity, it is most likely that they all correspond to the same face and are thus integrated into a robust single representation of the face in memory. Wallis and Bülthoff (2001) illustrated this with a rather extreme example, in which they observed that participants learned to associate a view of a completely different identity with the same face when all images were presented in spatiotemporal coherence. They presented four views of a face in smooth rotation and switched the identity of an intermediate view and observed that participants were likely to regard all face-views as belonging to the one person, having a direct effect on their memory of that face. Yet when the view sequence was ‘jumbled’, disrupting spatiotemporal association, this effect disappeared. Spatiotemporal association has been suggested to be more relevant for the recognition of faces over objects and may in part reflect the nature in which we encounter faces in movement in the natural environment (Liu, 2007).

*Audio-visual person recognition*

Although the voice, like the face, is uniquely associated with a person’s identity, most research has focused on how we process vocal information for the purpose of perceiving speech, and it is only in recent years that attention has turned to how we process paralinguistic information in speech to support speaker recognition. Yet in evolutionary terms speech is only a recent addition, and our ancestors have been deducing information from vocalizations to support recognition for millions of years (Belin et al., 2004). The fundamental nature of voice recognition is apparent not just from birth, with newborns showing a preference for their mother’s voice (DeCasper & Fifer, 1980), but there is evidence to suggest that this ability to discriminate voices may in fact emerge in utero (Fifer & Moon, 1994).

The voice, like the face, is a unique identity marker and conveys similar identity information including age, gender, ethnicity and familiarity as the face and has thus been referred to as an ‘auditory face’ (Belin et al., 2004). Indeed, the perceptual system is tasked with a similar problem in decoding identity in either the voice or the face. Like faces, voices share the same basic perceptual features across individuals (acoustic parameters such as pitch and timbre), and thus the perceptual system must represent a unique ‘vocal signature’ for each human voice in memory, by perceiving subtle differences in these acoustic parameters (Belin, Bestelmeyer, Latinus, & Watson, 2011; Belin et al., 2004). Although our ability to recognise voices is not as efficient as our ability to recognise faces (Hanley, Smith, & Had, 1998; Stevenage et al., 2012; Stevenage, Howland, & Tippelt, 2011), we can deduce identity information from the voice alone (Kreiman & Papcun, 1991; Papcun, Kreiman, & Davis, 1989; Schweinberger, Herholz, & Sommer, 1997).

Indeed, the neocortex has developed a network that supports voice recognition that is structured along the same principles as the network that supports face recognition. As the FFA has been implicated in the recognition of facial identity (Grill-Spector et al., 2004; Kanwisher et al., 1997; Parvizi et al., 2012),
recent findings have revealed voice selective regions, specifically temporal voice areas (TVA), which are located in mid and anterior regions of the superior temporal sulcus (mSTS and aSTS, respectively) (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin et al., 2004; Belin, Zatorre, & Ahad, 2002). These areas are sensitive to, among other things, identity information in speech. Similar to the prosopagnosic literature, insult to these areas, particularly when localised to the right hemisphere, can result in deficits in voice recognition known as phonagnosia (Van Lancker & Canter, 1982) and recent evidence also suggests the existence of developmental phonagnosia (Garrido et al., 2009). In contrast, damage to left hemisphere areas often result in deficits in speech perception (aphasia) but with intact voice recognition (Van Lancker & Canter, 1982).

**Integrating the face and voice for person recognition**

It appears that the brain has evolved to efficiently process face and voice information in a similar fashion. Recent evidence suggests that vocal and facial information is integrated and shared across modalities at earlier stages of processing (Ghazanfar, 2011; Ghazanfar & Schroeder, 2006) than were previously assumed (Burton, Bruce, & Johnston, 1990). Pioneering work by von Kriegstein and colleagues reported intriguing findings that the FFA is active during familiar voice recognition, even when there is no visual input (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005). This adds to the growing evidence that activation in unisensory areas of the brain can be modulated by cross modal input (Alais, Newell, & Mamassian, 2010; Driver & Spence, 2000; Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006) and suggests that the ventral temporal lobe may represent such person identity information *independent* of the input modality. Indeed, it has been reported that the even the haptic and the visual modalities may encode face information in a similar way, for the purpose of supporting face recognition (Casey & Newell, 2007).

In addition, von Kriegstein et al. (2005) observed that when the task of the participant was to focus on the identity of the voice, activation between the FFA and TVA was apparent. As mentioned such areas are involved in the processing of identity information from the voice (Belin et al., 2011). Such direct sharing of information is evolutionary advantageous, as it allows for robust person recognition, through cross-modal compensation, under both suboptimal viewing and/or listening conditions. This may be evident with decline in the sensory modalities, as occurs with the ageing process (Maguinness, Setti, Burke, Kenny, & Newell, 2011) or in a dimly lit or noisy room where it may be difficult to see a person’s face or to hear the person’s voice.

Research reported by von Kriegstein and Giraud (2006) has provided further evidence for the existence of an audio-visual person recognition network. They demonstrated, in an fMRI study that the learning of audio-visual face and voice pairings increased the functional coupling between face and voice areas in the cortex. Furthermore this functional coupling was observed only for natural audio-visual pairings that offer some level of cross-modal redundancy (e.g. a face and a voice versus a phone and a ring tone). Blank, Anwander and von Kriegstein (2011) conducted a follow up study using diffusion tensor imaging (DTI). DTI offers the possibility of assessing the connectivity of
brain areas through probabilistic tractography. Blank et al. (2011) demonstrated direct structural connections between the FFA and the voice identity sensitive areas in aSTS and mSTS. Such direct sharing of information between regions may be used to optimize person recognition. Thus, at least for the naturalistic coupling of face and voice stimuli, the multisensory effects may result from direct interactions between voice and face recognitions areas in the brain, as opposed to top-down modulation from supra-modal areas, as was once thought.

In line with the findings reported by von Kriegstein and colleagues, Schweinberger, Kloth and Robertson (2011) revealed evidence for face and voice interactions occurring at strikingly early stages of processing. They demonstrated through electroencephalogram (EEG) recordings, that event related potential (ERP) responses occur in the frontocentral areas at around 50-80 milliseconds (ms) following audio-visual (AV) presentation. These ERP responses occurred earlier and were more pronounced (super-additive) for AV presentations, compared to visual or auditory unimodal presentations. Moreover they also noted that AV presentations induced a larger N170 response than visual only presentations of faces (Schweinberger et al., 2011). The N170 response is associated with, and consistently evoked by, the visual presentation of a face, and is thus thought to be a key marker of processing that is specific to faces. Schweinberger et al. (2011) argued that such early audio-visual processing is mandatory; however it may not reflect audio-visual integration for the purpose of recognition. For example, detecting AV incongruencies in person recognition emerges at a later stage in the processing stream: mismatches in face and voice identity elicits differences in ERP responses at around 250-600 ms (Schweinberger et al., 2011). However, it is important to note that this is most likely the optimal way for person perception to have evolved, when we consider that is unlikely that we would ever encounter a face and a voice articulating speech in spatial and temporal alignment that did not belong to the same identity. As such these incongruency effects tell a lot about the optimal nature of the processing ability.

The neurological effects observed above have also been shown to have direct behavioural consequences. For example, Schweinberger, Robertson, and Kaufmann (2007) presented individuals with dynamic face and voice pairings, preserving the cross-modal spatial and temporal alignment of the faces and voices, while manipulating the congruency of identity in these face and voice pairings (such that a voice could be presented with a non-corresponding facial identity but the face would appear to be articulating with the voice in synchrony). They noted behavioural benefits and costs for identifying familiar voices, when they were respectively accompanied with either a congruent or incongruent facial identity. Yet importantly this was particularly true for voices accompanied with time synchronised dynamic facial information, static facial information either reduced or removed the effects of cross modal modulation, demonstrating that spatial and temporal alignment is crucial for audio-visual integration in person recognition. Indeed dynamic information assists in recognising individuals not just in audio-visual conditions, but also across modalities and we are efficient at matching a voice to a
moving face even when the utterance content is completely different (Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Lander, Hill, Kamachi, & Vatikiotis-Bateson, 2007; Munhall & Buchan, 2004).

Research from von Kriegstein and Giraud (2006) suggests how audio-visual information is also beneficial in creating a robust percept of a person in memory. They observed increased accuracy for recognizing voices that had been originally paired in the learning phase with their associative faces. Similarly O’Mahony and Newell (2012) observed that participants who were trained on face and voice pairings were faster in detecting the familiarity of a person when they were presented with congruent face and voice information at the testing stage. Intriguingly they did not observe the same recognition enhancement for congruent face and name pairings, suggesting that it is the naturalistic coupling of face and voice information which enhances person recognition and not the pairing of the face with additional arbitrary information i.e. the name of the person.

IN SUMMARY

Person recognition underpins many other cognitive tasks, including our ability to socially interact with the world. The fundamental nature of this skill is highlighted by the fact that person recognition appears to be a unique cognitive ability, which may be genetically determined. However, person recognition is adapted to and tailored towards the pressures of our surrounding environment and is malleable to change throughout development, even into old age. This interaction of genetics and environment has ultimately resulted in our high level of expertise in recognising others, which in turn is likely to have mediated our survival throughout our evolutionary history.

Acknowledgements

Face stimuli in Figure 2 are taken from the Radboud Faces Database (RaFD) (Langner et al., 2010) and are available for download, with permission, from http://www.socsci.ru.nl:8180/RaFD2/RaFD?p=main
Figure 1. Example of average face images, created using a composite of 100 individual male faces. Each image reflects an average face for males in their mid-20s (left) and mid-40s (right). Typically both images will be deemed more attractive than their individual counterparts.

Figure 2. Although face-images 1 and 2 share more visual similarity than faces 1 and 3, the perceptual system must become tuned to recognising the face of the same individual over large changes in visual appearance and therefore recognise faces 1 and 3 as the same facial identity. (Images from Langner et al., 2010).

Figure 3. Infants prefer to look at visual stimuli which schematically represent a face (left), versus those that do not (right). Stimuli adapted from (Valenza et al., 1996).
References


Maginness & Newell


