Investigating the structure of person perception and its broader affective and cognitive correlates: an individual differences approach

Hannah Louise Connolly

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology

Royal Holloway, University of London

2019

Declaration of Authorship

I, Hannah Louise Connolly, hereby declare that this work, which is approximately 56,000 words in length (excluding references), was carried out in accordance with the Regulations of the University of London. I declare that this submission is my own work, and to the best of my knowledge does not represent the work of others, published or unpublished, except where duly acknowledged in the text. No part of this thesis has been submitted for a higher degree at another university or institution.

Signed

Date: 27/11/19

Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Regulations appropriate for the degree of PhD at the University of London and that the candidate is qualified to submit this thesis in application for that degree.

Signature of supervisor:

Date: 27/11/19

For Valerie

Acknowledgements

First and foremost, huge thanks go to my supervisor, Dr Gary Lewis. Gary – I can't thank you enough for all the guidance and support you've given me over the last four years. You showed me what it is to do good science, to do work you're proud of, and to be persistent and resilient. Thanks for taking a chance on me at York and letting me come and work with you. The lessons you've taught me have been (and will continue to be) absolutely invaluable.

Secondly, to the rest of the Royal Holloway PhD students who made the academic office environment a place of fun and liveliness. Special thanks go to Lilla who has become a very close and valued friend over this doctoral experience. Thanks also to Rachael and Isaac for making the Lewis Lab an entertaining and thought-provoking group of which I was very glad to be a part.

Huge thanks go to my collaborators, Carmen Lefevre and Andy Young, without whom the publications from this doctoral work would not have been possible. Their thorough comments undoubtedly made my work much stronger. Andy in particular inspired the start of my research into face perception during his lectures and modules at the University of York, and I feel humbled to have received his expert input and ideas.

Finally, a big thank you to my wonderful family for their ongoing love and encouragement. To my parents, Veronica and Stephen (the original Dr Connolly) - thanks for teaching me to work hard and to trust that the dots will always join up. To Rebecca (and Angus) – thank you for being there for me, and for your help and belief throughout. To Jon, for making me smile every single day. This thesis is for all of you.

Abstract

The ability to perceive social information from another person is an important skill for successful everyday functioning. Given that deficits in this ability can be detrimental for wellbeing and relationships, it has unsurprisingly been the focus of much research. This has revealed notable individual differences in the general population; but relatively little work exists seeking to understand this variation. Moreover, sex and age differences have largely been studied with small samples and problematic stimuli. In this thesis, I addressed a number of major issues in the field of person perception through the lens of individual differences.

In Chapter 3, I examined the individual differences architecture of emotion recognition across three modalities (face, body, voice), and observed that this ability is related to, yet distinct from, the recognition of facial identity, the construct of alexithymia (a trait reflecting difficulty in identifying and describing feelings), and general intelligence. In Chapter 4, I observed a moderate association between emotional expression and identity recognition even after adjusting for general intelligence. In Chapter 5, I observed a negative relationship between alexithymia and intelligence, independent of emotion recognition ability.

In Chapter 6, I examined sex differences in facial and bodily emotion recognition, and observed a small but consistent female advantage for recognising facial disgust. Finally, in Chapter 7, I examined the effect of age on facial expression and facial identity recognition, and report significant age-related declines in both abilities that were independent of each other and of general intelligence.

This thesis expands our understanding of key components of person perception. I have outlined a proposed structure of emotion recognition ability, and how this is related to identity recognition, cognitive ability, and demographic variables. I suggest that individual differences are highly important to consider at both the general (cognition) and specific (face-processing) levels of this ability.

Contents

Bibliography of Work Completed During PhD

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2019). Sex differences in emotion recognition: Evidence for a small overall female superiority on facial disgust. *Emotion, 19*(3), 455-464.

Connolly, H. L., Young, A. W., & Lewis, G. J. (2019). Recognition of facial expression and identity in part reflects a common ability, independent of general intelligence and visual short-term memory. *Cognition and Emotion, 33*(6), 1119-1128.

Connolly, H.L, Lefevre, C.E, Young, A.W, & Lewis, G. J. (2020). Emotion recognition ability: evidence for a supramodal factor and its links to social cognition. *Cognition*, *197*, 104166.

Connolly, H. L., Young, A. W., & Lewis, G. J. (under review). Face perception across the adult lifespan: Evidence for age-related changes independent of general intelligence. *Cognition and Emotion.*

Connolly, H. L., Young, A. W., & Lewis, G. J. (revision submitted). Consistent Evidence of a Link between Alexithymia and General Intelligence. *Cognition and Emotion.*

Chapter 1: Introduction

Person perception involves attending to social information from another individual, and being able to comprehend and integrate this information to give the most appropriate response. This ability is highly important for forming and maintaining successful relationships with conspecifics, and deficits in person perception can result in psychiatric difficulties and social isolation. In this thesis, I will focus on two main aspects of person perception: recognition of emotional expressions, and recognition of facial identity. Both of these abilities are necessary for everyday functioning, and as such, have been the focus of substantial research. However, as I will outline further, there exist several limitations in the way that these abilities have been assessed, and this thesis seeks to address some of the major concerns.

The first aspect of person perception, emotional expression recognition, is fundamental for nonverbal communication. It enables individuals to understand other people's actions and infer their intentions, which then allows them to respond most appropriately (Bal et al., 2010). Emotional expressions can signal a source of threat in the environment, and as such, fast and accurate perception is highly adaptive for human survival. Furthermore, successful emotion recognition is associated with better social functioning (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006) and greater relationship well-being (Carton, Kessler, & Pape, 1999). Conversely, an impairment in emotion recognition is associated with psychopathological disorders including (but not limited to) depression (Surguladze et al., 2004), social anxiety (Joormann & Gotlib, 2006), and borderline personality disorder (Fenske et al., 2015).

The second aspect of person perception I will focus on, face identity recognition, is also necessary for everyday interactions. Being able to instantly recognise a familiar person's identity from their facial features allows one to adjust one's response based on prior experiences with and knowledge of the other person. A deficit in this ability can be socially debilitating and impact negatively on one's quality of life, social interactions, and mental health (Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008).

For emotional expression recognition and particularly identity recognition, much of the past research has tended to assume that the neurotypical population are similarly competent in their recognition ability (Rozin, Taylor, Ross, Bennett, & Hejmadi, 2005). The early literature also suggests that it is mostly individuals with specific neurological or psychiatric conditions whom exhibit difficulties with these everyday abilities, and traditional models have supported this dichotomous view. Some of the main theories are outlined below.

1.1 Insights from Cognitive Models of Face Processing

There have been two key cognitive models proposed that outline how the brain may process diverse information from the face. In their functional model of face recognition, Bruce and Young (1986) argued for a clear distinction between the processes underlying emotional expression and identity recognition. The model initially posited separate pathways for emotional expression and identity recognition that diverged after an early stage in face perception in part because of clinical studies of individuals with brain lesions, in which double dissociations between expression and identity were observed. Studies of brain lesion patients showed selective deficits on recognition facial expressions and facial identity (Young, Newcombe, de Haan, Small, & Hay, 1993). Some separation between these processes makes sense given we do not need to be familiar with an individual in order to decode their facial expressions, and we can easily recognise a familiar face regardless of the expression they are making.

In line with Bruce and Young's model of functional separation for these two abilities, Haxby and colleagues' cognitive neuroscience model (Haxby, Hoffman, & Gobbini, 2000) posited distinct brain regions for processing of variant and invariant facial cues. Whilst invariant, stable characteristics (identity, age, sex) are processed in ventral occipitotemporal regions such as the lateral fusiform gyrus, variant facial cues (eye gaze or facial expression) involve the superior temporal sulcus (STS). These findings from neuroimaging studies of separate routes for expression and identity processing neatly reflect the functional distinction in Bruce and Young's model, and therefore the framework was bolstered by converging evidence from different research areas.

More recent work, however, has suggested a more nuanced association between these two core processes of face perception. Neuroimaging studies using fMRI have examined brain areas that were previously considered to be specialised to either facial expression (posterior STS) or facial identity (fusiform face area), and demonstrated that these regions can be activated in response to changes in either type of facial signal (Fox, Moon, Iaria, & Barton, 2009). Additionally, an asymmetric relationship has been reported in which variations in facial identity influenced the perception of facial expression, but variation in expression did not affect perception of identity (Schweinberger & Soukup, 1998). This suggests a greater level of dependence between expression and identity recognition that was not predicted by the Bruce and Young or Haxby et al. models.

It also suggests that the structure of face perception may reflect the demands of daily reality, insofar as it being necessary to recognise different facial expressions on the same person without perceiving a change in identity. Moreover, it seems reasonable that changes in identity can affect judgements of expression. Several studies have reported that changes in invariant facial signals, for example, face shape or sexually dimorphic facial features, can influence how variant facial signals (expressions) are perceived. As an example, Sacco and Hugenberg (2009) demonstrated that making faces appear more baby-like, i.e. by making them rounder with larger eyes, facilitated the perception of a fearful facial expression, whilst making the faces appear older (smaller eyes and narrow face shape) facilitated the perception of anger.

Other studies have noted that the judgements on sex and facial expression of a face are not entirely independent of each other. Rather, they show an interaction effect in that participants are quicker and more accurate at detecting an angry male face and a happy female face (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007). From this, it seems possible that there is some natural overlap in the features that denote masculinity and anger, e.g., heavy brow, inset eyes, large jaw, and also some overlap in features denoting femininity and happiness, e.g., rounded cheeks (Becker et al., 2007). What is clear from these studies is that processing of variant and invariant facial features is not as functionally distinct as classic theory would suggest.

1.1.1 Individual Differences in Face Identity Recognition

Whilst more recent work has indicated that processing of variant and invariant facial signals is not as functionally distinct as posited by traditional cognitive models, it still remains unclear as to the specific degree of overlap between these processes. Regarding facial expression and identity recognition, recent perspectives suggest that the evidence for overlap might arise from the early common perceptual stage that both abilities share, and that the evidence for functional separation comes from after the pathways diverge later in the face processing system. There exists some evidence for this, although more work is needed to clarify the extent of the overlap.

Importantly, the widely accepted traditional view has been that the majority of the typical population perform similarly well on measures of face processing, and only individuals with neurological impairment will show relative difficulties. Recent work, however, has identified a greater variation in ability across the entire population than was typically assumed (e.g. Burton, White, & McNeill, 2010). An individual differences perspective, then, is a highly worthwhile avenue to offer further insight on this issue, and the work of this thesis seeks to contribute to the debate.

1.2 Insights from Evolutionary Models of Emotion Recognition

Following on from classic cognitive models of face processing, a plausible explanation for the ability to recognise the variant signals of emotional expressions in others is that such ability evolved in order to serve an adaptive function. This evolutionary perspective was first outlined in Darwin’s seminal work on facial expressions of emotion (1872/1965) and inspired a great deal of academic debate in the field as to whether certain facial expressions are innate and universal, as posited by Darwin, or the degree to which they are culturally dependent, as proposed by others. Many researchers since Darwin’s work have argued in favour of a small subset of emotions that are termed ‘basic’, meaning fundamental and distinguishable emotional categories that may combine to form more complex emotions (Ekman, 1999). Nativist theories of emotion posit discrete emotional categories that are innate, adaptive, and initiated automatically and unconsciously by a natural, evolutionarily-driven stressor in the environment (Izard, 2007). However, much discussion still centres on which emotional categories should be included in this ‘special’ status, as well as the biological or evolutionary mechanisms for why they are considered basic.

Arguably the most common reason to posit the existence of basic emotions is to explain the empirical findings that some emotions seem to exist across all cultures. In a series of studies, Ekman and colleagues examined various developed Western and Eastern cultures (e.g. Brazil, United States of America, Argentina, Chile, and Japan), and found comparable identification of emotional expressions across all samples. Importantly, this finding extended to two pre-literate cultures from Borneo and New Guinea (Ekman, Sorenson, & Friesen, 1969), thus giving support to Ekman's proposal that certain facial movements are universally associated with certain emotional categories. Ekman argued that these findings argue against the idea that culture-specific norms solely underpin the expression and recognition of basic facial emotions. Rather, cultural differences may instead emerge in the display norms associated with culture-specific social settings. Being able to adapt our human biological propensity to unique cultural environments is an adaptive and flexible faculty that has likely arisen from evolution (Nesse, 1990).

If, as Darwin, Ekman, and others argue, the basic emotional expressions are intrinsic and universal, then an interesting question follows as to how these expressions have come about. Some emotion researchers have suggested that the basic emotions are biologically primitive and 'hard-wired' into our biological make-up. That is, there are features in the basic emotion categories that make them adaptive in an evolutionary sense, both in a non-communicative, physiological way, but also as a means to quickly and effectively communicate important information to conspecifics. Successful expression and recognition of these emotions, then, is required for future survival and reproduction. Expressing disgust, for example, consists of wrinkling one's nose, narrowing the eyes, and expelling the tongue, all actions which minimise the inhalation or congestion of potential contaminants in the environment (Rozin, Lowery, & Ebert, 1994). Being able to rapidly and accurately recognise these signals of disgust in a conspecific may alert the individual to harmful odours or a putrid foodstuff, and lead them (or their offspring) to avoid ingesting these potential toxins. This argument also holds for recognition of fear. The facial (and subsequent vocal) expressions of fear are automatic, high-contrast signals of immediate threat in the environment. Being able to recognise these signals at distance serves as a warning of danger that enables the individual and other members of the group to flee and reach safety or prepare for attack.

Finally, whilst outbursts of anger can often be viewed as irrational, the basic emotion still may have an adaptive basis. Angry expressions are associated with a narrowing of the eyes and lowering of the brow, features that are thought to be in preparation for an attack, as well as to signal the individual's physical strength. These features of an angry facial expression may have in part evolved in order to signal physical dominance (lower brow, large jaw) and maturity (small eyes) to deter any would-be attackers in the group (Sacco & Hugenberg, 2009). Being able to recognise when someone (with greater physical strength) is angered and considering an attack would be hugely important for preparing for or attempting to avert the attack.

1.2.1 Individual Differences in Emotional Expression Recognition

As already outlined in the context of face identity recognition, traditional cognitive models have suggested that the majority of the population have similar recognition ability, but it is clear that individual differences have largely been overlooked. To a lesser extent, the same could be said for emotional expression recognition. Evolutionary theories of how emotions have evolved to serve an adaptive purpose suggest that neurologically healthy people are similarly skilful in their emotion recognition, but notable individual differences have been reported in this domain, for example, Lewis, Lefevre, and Young (2016). Studies such as Lewis et al. with sensitive measures have demonstrated that there is substantial variation in what is considered 'typical' recognition, with some individuals performing very accurately and others performing poorly.

Many earlier emotion recognition studies used posed prototypical stimuli in which the target emotion was easy to identify and thus often resulted in ceiling effects. In addition, some emotional expression stimulus sets were developed to distinguish between the neurotypical population and people with neurological deficits. These stimulus sets are therefore not suitable for examining the natural variation in the neurotypical population. This thesis seeks to overcome some of these limitations and use stimuli that have been morphed to different intensities and carefully piloted. This ensures their suitability for detecting subtle individual differences in performance, and extends our understanding of this ability in the general population.

1.3 Individual Differences in Intelligence

A further variable to consider when examining individual differences in person perception is that of intelligence, given the known large variation that exists in the general population. To the extent that facial expression and facial identity recognition do share common perceptual processes, an interesting question is if this common stage of overlap is reflective of broader cognitive processes. There have been mixed findings concerning the association of general cognition to face perception abilities. In the domain of facial expression recognition, several studies have noted a moderate positive correlation with general intelligence (Borod et al., 2000; Lewis et al., 2016; Schlegel & Scherer, 2016), although a significant association has not been noted unanimously (e.g. Palermo, O'Connor, Davis, Irons, & McKone, 2013).

In the domain of facial identity recognition, an absence of a significant relationship with intelligence has often been reported (e.g. Palermo et al., 2013, Wilmer, Germine, & Nakayama, 2014), and identity recognition has been considered a 'special ability' distinct from broader cognitive abilities (Wilmer et al., 2010). However, other studies do report a modest association with general intelligence (e.g. Shakeshaft & Plomin, 2015), and as with many other areas of research, the discrepancy may have arisen from differences in methodology, samples, and measures. This thesis seeks to disentangle the individual differences that exist in both emotion and identity recognition from the independent differences present in general intelligence.

1.4 Individual Differences in Emotional Expression Recognition: Factor Structure

As outlined above, person perception abilities have been a topic of enduring interest. Regarding the individual differences *structure* of this ability, however, there exists far fewer studies. Of those, there is debate as to whether all facial expressions are processed by a single, general facial affect recognition system or whether specific neural networks underlie recognition of discrete categories of individual emotions. This section will first consider evidence from both behavioural and neuroimaging studies in order to examine the main theoretical perspectives in the field. Secondly, it will consider individual differences work that has used a factor analytic approach, to give an outline of the proposed structure of emotion recognition.

1.4.1 Emotion specificity: brain injury and neuroimaging studies

To date, arguments for functional specialisation of emotion recognition have in part arisen through neuropsychological work involving participants with brain injury. These studies have suggested that specialised neural networks are implicated in the recognition of individual emotions, and that damage to these areas results in a deficit in recognising a particular emotion. For example, previous work has observed that patients with amygdala lesions often fail to recognise fearful facial expressions (Adolphs, Tranel, Damasio, & Damasio, 1994).

In contrast, other researchers have argued that amygdala-damaged individuals can exhibit difficulties with other emotions as well, for example, disgust (Rapcsak et al., 2000). Moreover, Rapcsak and colleagues reported no significant difference in the recognition of fear between subjects with or without amygdala damage, evidence against a selective fear impairment in the amygdala-damaged population. Thus whilst damage to various brain structures can disrupt typical emotion recognition, this disruption may not necessarily affect specific emotions, even if specialised neural networks are damaged (Rapcsak et al., 2000).

Further empirical support for the emotion-specific networks hypothesis has been argued in the domains of other emotions. Firstly, damage to the basal ganglia, for example in Huntington's disease or Obsessive Compulsive Disorder, has been associated with impaired recognition of disgust (Sprengelmeyer et al., 1997). Furthermore, healthy participants given a dose of anti-anxiety drug Diazepam showed a selective impairment in recognition of anger whilst the other emotions were unaffected (Blair & Curran, 1999).

1.4.2 Valence specificity

Whilst a number of studies have argued for specialisation of recognition of individual emotions, other work has argued instead for differential recognition of positive and negative emotions. Adolphs and colleagues (1999) argue that although amygdala damage has been shown to impair fear recognition in particular, it also impairs recognition of the other negative emotions, whereas happiness recognition remains preserved. In contrast, meta-analytic work by Ruffman and colleagues (2008) reported that age-related decline largely affects recognition of the negative emotions, but that it did have some impact on happiness recognition. They also reported that older groups showed relative preservation of disgust recognition, which neurological perspectives have argued may be due to structures within the basal ganglia showing a slower age-related decline than areas associated with other emotions, for example, the amygdala (Ruffman, Henry, Livingstone, & Phillips, 2008).

The inconsistency in the reported valence effect findings may have arisen, at least in part, due to inequalities in the tested emotions themselves. Many studies test four negative emotions (anger, disgust, fear, and sadness) and two positive emotions (happiness and surprise), the last of which is debated as to its inherent valence (Rosenberg, Dethier, Kessels, Westbrook, & McDonald, 2015). Additionally, research indicates that these six emotions are not equally difficult to recognise, for example, average accuracy scores are reported as highest for happiness (94%), and lowest for fear (70%), and that putative valence effects may be confounded by ceiling effects in the positive emotions.

1.4.3 Emotion generality: evidence from behavioural studies

In the behavioural domain, studies suggest a general ability in recognising emotions across categories, and also extend this to recognition of emotions across different sensory modalities. In a study of 166 individuals, Rozin and colleagues (2005) tested the four negative emotions (anger, disgust, fear and sadness) in four modalities (face, body, dance and hand gestures), and noted significant positive correlations between the individual emotions (.42-.62), and also between the modality subtests (.29-.54). Whilst there was considerable variation in overall accuracy scores within the sample, there was no evidence of participants showing a selective difficulty with any particular emotion (Rozin et al., 2005).

Behavioural evidence also suggests a substantial overlap in recognition of emotions between the visual and auditory modalities, at least when presented simultaneously. de Gelder and Vroomen (2000) asked participants (N=44) to interpret the emotional state of either a static face or a spoken sentence. Even when instructed to ignore one of the sensory sources, participants were still affected by this concurrent information when making the emotion judgement, suggesting that the audio-visual integration of emotional information is somewhat automatic and involuntary. It also appears to be advantageous: a robust behavioural finding in the literature is that unimodal (i.e. face only) emotion displays are recognised at a slower and less accurate level than the congruent bimodal (i.e. face and voice) condition (Dolan, Morris, & de Gelder, 2001).

1.4.4 Emotion generality: evidence from neuroimaging studies

Neuroimaging studies also offer compelling evidence of emotion recognition across modalities. A range of brain regions show greater activation to multimodal than to unimodal emotion expressions. Specifically, the left posterior superior temporal sulcus (pSTS) meets Calvert, Campbell, and Brammer's (2000) criteria of a multimodal region, by demonstrating a supra-additive response to audio-visual stimuli (Hagan et al., 2009). Furthermore, transcranial magnetic stimulation (TMS) studies have shown that disruption to left pSTS activity results in reduction or loss of the McGurk effect (Beauchamp, Nath, & Pasalar, 2010). Calder and Young (2005) also posit parts of the pSTS as having cells that receive input from multi-sensory brain regions, and are involved in dealing with changeable aspects of faces, for example, lip movement and eye-gaze. It is thought that this region is crucial in integrating signals from different sensory channels in order to determine the emotional state and attention of conspecifics (Perrett, Hietanen, Oram, & Benson, 1992), as well as detecting incongruity in an individual's expressions that may suggest deception (Schirmer & Adolphs, 2017).

1.4.5 Emotion generality: evidence from structural equation modeling (SEM)

Suzuki, Hoshino, and Shigemasu (2010) (N_{total}=805) tested participants on the five basic emotions (anger, disgust, fear, happiness, and sadness) and found support for a general recognition ability across emotions. Their SEM analysis indicated that the best fitting model was a single factor that loaded substantially onto the four negative emotions (coefficients of .64) with a weaker loading onto happiness (.18), although all loadings were statistically significant. The authors suggested a minimal overlap between positive and negative emotions but noted that the valences were not completely independent from each other.

Schlegel, Grandjean, and Scherer (2012) (N_1=305; N_2=295) also reported that a general factor model fitted their data best with added minor factors for confused emotions, e.g. anger and irritation. Furthermore, a model positing a single factor underlying emotion recognition across the both audio and video presentation showed an excellent fit to the data, thereby supporting earlier behavioural and neuroimaging work that emotion recognition reflects a multimodal ability.

Schlegel and colleagues' (2012) second study also supported a general factor model (with additional latent factors for positive and negative emotions in which surprise correlated with both valence factors) showing the best fit to the data. Moreover, the high correlation (.77) between the two valence factors suggests a substantial overlap between positive and negative emotion recognition ability. The authors proposed that these two valence factors are sub-dimensions of a broad general emotion recognition ability, with valence factors being highly correlated, not distinct.

In the visual domain, Lewis et al. (2016) carried out three studies of face and body emotion recognition, and reported the best fitting model to be a higher-order structure in which there was an emotion-general multimodal factor acting indirectly on the face and body manifest variables, as well as direct face-specific and emotion-specific modality-general factors. These findings suggest that emotion recognition ability operates at broader emotion-general and modality-general levels, with more specific factors for individual emotions or modalities being subsumed under this single dimension.

In sum, across a range of behavioural (de Gelder & Vroomen, 2000), neuroimaging (Calder & Young, 2005), and structural modeling (Lewis et al., 2016) studies, the majority of evidence appears to suggest that a single underlying dimension best reflects the structure of emotion recognition, regardless of the sensory modality or emotion category. This thesis

(Chapter 3) seeks to contribute to the debate in this area by examining the individual differences structure of emotion recognition ability across visual (face, body) and auditory (voice) modalities. Importantly, I will assess this ability using stimuli that are sufficiently sensitive to detecting subtle individual variation in this domain. The association of this ability to clinical and demographic traits will also be examined. This will contribute to our understanding of the nature of emotion recognition ability, and place it in the context of wider person perception.

1.5 Outcomes of Individual Differences in Emotional Expression Recognition: Psychopathology

Investigating individual differences in person perception is of particular importance given its real-life implications. For example, there are a number of outcomes that are related to the ability to recognise emotional expressions. One of the most salient outcomes is the positive association between emotion recognition deficits and risk of psychopathology, particularly those disorders characterised by impairments in interpersonal functioning. A general impairment in processing emotional expressions has been noted for several disorders, including psychopathy, autism spectrum disorder, schizophrenia, depression, and social anxiety disorder.

Importantly for the focus of this thesis, individual differences exist in many of these dimensional disorders, both above and below clinical thresholds, but this has not been well examined. In the following section, I review the work to date and consider how an individual differences approach is necessary for furthering our knowledge of emotion recognition deficits in these heterogeneous populations.

1.5.1 Autism

Autism, or autism spectrum disorder (hereafter *autism*), is the collective term for the pervasive developmental disorders that are characterised largely by repetitive and rigid behaviours, difficulties with reciprocal social interaction, and impaired verbal and non-verbal communicative skills. These social difficulties can be apparent from late infancy onwards (Ozonoff, Heung, Byrd, Hansen, & Hertz-Picciotto, 2008), and include, although not limited to, lack of orientation to faces, impairment in the social use of eye gaze, difficulty initiating communicative interactions, and impaired social cognition (Weiss & Harris, 2001; Rao, Beidel, & Murray, 2008).

In particular, difficulty in the recognition of emotional expressions is often suggested as a core feature of the social impairment associated with these disorders. Many studies report generalised deficits across a range of emotions (Bal et al., 2010), whilst others report selective impairments for specific emotions such as fear (Humphreys, Minshew, Leonard, & Behrmann, 2007).

One important account recently proposed for the large discrepancy in emotion recognition in autism is that of co-occurring alexithymia (Cook, Brewer, Shah, & Bird, 2013). Alexithymia, outlined in more depth in the following section, consists of difficulties identifying and verbalising emotions, impaired interoceptive abilities, and a concrete style of thinking. The condition has a prevalence of 10% in the general population, but some degree of alexithymia exists in approximately 50% of individuals with autism (Cook et al., 2013), with some estimates as high as 65% (Bird & Cook, 2013). However, despite the high levels of comorbidity, these are independent constructs: not all autistic individuals are alexithymic and equally, some individuals experience severe alexithymia without evidence of autism.

This recent perspective suggested is that it is the comorbid alexithymic component in autism that is associated with their emotional impairments, rather than a deficit caused by autism itself (Cook et al., 2013). The authors observed that there were no group differences in expression recognition between the autistic sample and alexithymia-matched healthy controls. Alexithymia correlated with accuracy of facial expression recognition, and importantly, it remained a significant predictor of recognition accuracy even when adjusting for intelligence, age, sex, or autism. In contrast, autism did not account for a significant proportion of variance in expression recognition. These results may help to disentangle the highly mixed findings of emotional deficits in autism, given that the majority of studies have not adjusted for this commonly co-occurring construct. It is plausible that the studies that observed emotional deficits in autism had a greater proportion of highly alexithymic individuals, particularly in view of the different prevalences of the construct in autistic and neurotypical populations. Future studies, then, should measure alexithymia in order to distinguish if any emotional difficulties appear to be explained by autism or comorbid alexithymia.

1.5.2 Alexithymia

As outlined above, alexithymia is a subclinical personality construct associated with difficulty identifying and describing one's own emotions, impairment in differentiating emotional arousal from bodily sensations, and an externally-oriented cognitive style. Alexithymia frequently co-occurs with many developmental and psychological disorders, including autism, eating disorders, anxiety, depression, and somatoform disorders (e.g. limb pain, stomach disturbance).

With regards to emotion recognition ability, behavioural studies have consistently reported that higher levels of alexithymia are associated with decreased emotion recognition accuracy (Lane, Sechrest, Reidel, Shapiro, & Kaszniak, 2000). Montebarocci, Surcinelli, Rossi, & Baldaro (2011) proposed that lower verbal ability was associated with alexithymia, and observed that the difference in emotion recognition performance between people with low and high alexithymia disappeared when adjusting for verbal intelligence.

The perspective of verbal difficulties as a potential mechanism for alexithymia has been proposed in a recent paper by Hobson and colleagues (2019). In their *alexithymia-language hypothesis,* they argue that early language abilities contribute to typical emotional understanding and development, and therefore infant or childhood language difficulties may give rise to emotional difficulties in later life (Hobson, Brewer, Catmur, & Bird, 2019). They support this perspective with studies of children with speech and language impairments, hearing impairment or deafness, and autism spectrum disorder, all of whom show early deficits in typical language development, and which is associated with later difficulties in regulating and recognising emotions, empathy, and broader social interactions.

Converse to Montebarocci et al's and Hobson et al.'s verbal difficulties perspectives, Lane et al. (1995) proposed that the negative correlation between alexithymia and emotion recognition accuracy is a result of both verbal *and* nonverbal difficulties. They tested 380 participants on the Perception of Affect Task (PAT) which comprises four verbal or nonverbal emotional tasks. Individuals higher in alexithymia showed reduced accuracy across the total PAT, regardless of the verbal or nonverbal nature of the task. The authors argued that alexithymia has been long defined as being a difficulty with putting emotions into words, but that the condition may be better conceptualised as a more general deficit in emotion processing. In rebuttal, Hobson et al. (2019) argue that a nonverbal emotion recognition task fails to acknowledge the role of language in early emotional development, thus why alexithymic individuals will struggle on emotional tasks even in the absence of linguistic demands. In sum, it remains difficult to tease apart the verbal and nonverbal abilities that are important in alexithymia, and this thesis aims to offer an additional perspective on this ongoing debate.

1.5.3 Social Anxiety Disorder

Social anxiety is a type of anxiety that is marked by avoidance of situations with potential negative evaluation, a tendency which may result in damage to their educational, occupational, and social lives. In terms of emotion recognition, some studies report evidence of enhanced accuracy for negatively valenced faces (Winton et al., 1995; Mogg, Philippot and Bradley, 2004), whilst others report avoidance to angry faces, which the authors suggest as a mechanism for not wanting to engage with the facial expression at all (Mansell, Clark, Ehlers, & Chen, 1999).

In contrast, Button and colleagues (2013) reported that socially anxious individuals may not have a problem accurately decoding a facial expression such as anger *per se* but show differences relative to healthy controls in their interpretation of how threatening they perceive this expression to be. They also highlighted that misinterpretations were most likely to occur when the individual was uncertain (i.e., for more subtle expressions), and that this effect is not specific to negative emotions but is instead reflective of a general response bias in expression decoding (Button, Lewis, Penton-Voak, & Munafò, 2013).

The nature of emotional difficulties in social anxiety is particularly relevant for the current thesis in light of recent work highlighting how aspects of this disorder can resemble core social deficits of autism (White, Bray, & Ollendick, 2012). Through factor analytic work, these authors highlighted a substantial overlap in the constructs of social difficulties in both of these disorders. Specifically, both autism and social anxiety measures captured a lack of enjoyment and avoidance of social situations, and a preference for aloneness. Whilst these two latent factors were statistically distinct, they show considerable overlap and a high prevalence of comorbidity (up to 57%: Kuusikko et al., 2008), and therefore White and colleagues (2012) suggested that both disorders should be measured when examining either one of these populations. This may be particularly appropriate for high-functioning individuals, in which self-reported social difficulties could reflect either socially anxious or autism-like traits.

Given the considerable prevalence of emotional difficulties in autism, alexithymia, and social anxiety, then, as well as the highlighted comorbidities between them, the current thesis (Chapter 3: Study 1) will assess individuals on all three of these traits. In this way, any emotion recognition deficits observed will not be misattributed to a co-occurring disorder.

1.6 Individual Differences in Person Perception

A further salient factor in person perception is how individual differences are influenced by demographic variables. In this section, I will firstly review the literature on how a person's sex interacts with emotion expression recognition ability, given the widely held view of female superiority in this domain. Secondly, I will review how both facial emotion and facial identity recognition are reported to change across the adult lifespan. In both, I will consider in particular how individual differences have been somewhat overlooked, and that using more suitable tests is crucial to detecting any subtle differences that may exist.

1.6.1 Sex Differences

Sex differences in emotional expression recognition have been examined extensively, but findings in this area have been mixed. A small but significant advantage for females is reported in the four meta-analyses that have been carried out (Hall, 1978, 1984; McClure, 2000; Thompson & Voyer, 2014). However, other studies report more mixed or null findings (e.g. Rahman, Wilson, & Abrahams, 2004; Lyusin & Ovsyannikova, 2016). This is due, at least in part, to the wide variety of stimuli and design that have been employed across the numerous studies. Many use a forced-choice paradigm (Rotter & Rotter, 1988; Hoffmann, Kessler, Eppel, Rukavina, & Traue, 2010) with a variety of tested basic emotions. Others have used a visual search task (Williams & Mattingley, 2006; Sawada et al., 2014). The stimuli chosen have typically been prototypical expressions of posed emotion, such as the Ekman set of faces (Ekman & Friesen, 1976), and the relative ease of recognition of these expressions often generates maximum accuracy scores and ceiling effects that obscure any group differences. Moreover, the sample sizes in many of these studies are very small (i.e. N<100) and therefore they lack the statistical power to detect any potential, and likely modest, effects. In addition, the most recent meta-analysis (Thompson & Voyer, 2014) indicated the presence of publication bias in the field, so the small general advantage for females in emotion recognition should be interpreted with some caution.

This thesis (Chapter 6) will address some of the challenges faced by previous work in this area, and attempt to offer more conclusive findings. Sex differences in both facial and bodily emotional expression recognition will be assessed using stimulus sets that have been morphed to different intensities to increase task difficulty and avoid ceiling effects. Any significant findings from the first sample of this chapter will be tested in two further independent samples so that results are more robust and replicable than have typically been reported in earlier studies.

1.6.2 Age Differences

Adult age is another important demographic variable that can contribute to the individual differences observed in person perception abilities. Much emotional development reaches peak performance by early adulthood, but an interesting question is how this changes with advancing adult age. That is, for how long does optimal performance last, and when and how does age start to impact on this ability? Studies examining age effects on emotion recognition have consistently reported declines across the basic emotions which is evident from middle adulthood onwards (Calder et al., 2003). These authors implicated the specific emotional categories of anger, fear and sadness, with disgust recognition showing evidence of a relative improvement or preservation with age.

A meta-analysis of 28 datasets (total N=1667) supported the general consensus of age-related difficulties in emotion recognition, which is evident across sensory modalities (voices, bodies, and face-voice matching) and basic emotional categories. In line with much of the previous research, particular difficulties were noted for anger and sadness and to a lesser extent, fear recognition, in older adults (average age: 71) relative to younger adults (average age: 24) (Ruffman et al., 2008). The authors also noted the preservation of disgust recognition that is observed in other studies.

A key question concerning age effects of emotion recognition is the mechanism responsible for the observed decline. As outlined above, one posited account is the known age-related declines in general intelligence (Deary, 2001; Salthouse, 2010) and/or in sensory abilities (Lima, Alves, Scott, & Castro, 2014). Emotion recognition performance may not be restricted by understanding and perception of emotional expressions specifically, but by broader task demands. A decline in fluid abilities such as processing speed, working memory, spatial reasoning, or impairment to vision and/or hearing may substantially contribute to the observed decline across emotion recognition tests.

Some studies have sought to include control variables of cognition, but have frequently used rough proxies, for example, level of education (e.g. Mill, Allik, Realo, & Valk, 2009; Sasson et al., 2010) or a single test of matrix reasoning or processing speed (e.g. Horning, Cornwell, & Davis., 2012; West et al., 2012). Accuracy did not change when adjusting for these various proxies of cognition, suggesting an independent effect of age on emotion recognition. However, a broader battery of cognitive tests is required to fully capture the variance associated with intelligence. The question of age effects on facial emotion recognition will be examined in this thesis (Chapter 7) using a well-established battery of general intelligence.

1.7 The Current Thesis

As outlined in the sections above, accurate perception of emotional expressions and facial identities is crucial for successfully navigating the social world and for being made rapidly aware of imminent threats in the physical environment. This thesis will examine a cluster of these person perception abilities, and use an individual differences approach to overcome some limitations of earlier work in the field.

Firstly, we know that emotions can be expressed through a variety of sensory channels, most commonly the face, body, and voice. Whilst we can recognise expressions from across these channels, little is known regarding the individual differences structure of emotional expression recognition ability. That is, if you are particularly good at recognising facial expressions of emotion, does this mean you are equally good at decoding bodily and vocal expressions as well? This research question comprises the first main aim of this thesis. In Chapter 3 (Study 1), I will use structural equation modeling to examine whether a general factor underpins recognition of expressions across these three sensory modalities.

If such a general factor appears to exist, the second main aim of the thesis is to determine how this general ability relates to broader non-emotional cognitive and affective traits (Chapter 3: Studies 2 and 3). This will help to clarify the specificity of the ability factor. In other words, is this putative factor associated only with expression recognition or does it encompass other face perception abilities or general intelligence? In Chapter 4, I will further investigate this question. Specifically, I aim to explore the degree of separation between facial expression and facial identity recognition, as posited by classic cognitive models of face perception (Bruce & Young, 1986; Haxby et al., 2000), and also whether any shared variance is reflective of broader cognitive or perceptual functioning.

Further to the question of broader links to expression recognition, the third main thesis aim is to investigate the relation of the emotion factor to general cognitive ability and to clinically-relevant traits involving emotional difficulties. In Chapter 5, I will examine whether the alexithymia construct shows an association with intelligence when adjusting for deficits in facial and supramodal (face, body, voice) emotion recognition ability.

The fourth main research aim examines how emotional expression recognition is affected by a person's sex. It has often been reported in earlier literature that females show superior performance in this regard, but empirical samples have typically been small and ceiling effects are common. Using robust samples of data, Chapter 6 will examine sex differences in facial and bodily expression recognition across three large, healthy samples, with a view to offering a more reliable conclusion than has previously been available.

Finally, the fifth research aim of the thesis concerns how person perception changes across the adult lifespan. In Chapter 7, I will leverage data from a large, age-diverse, healthy sample to examine how facial expression recognition and facial identity recognition change across adult age, and whether any change in these abilities is independent from the widely established age-related decline in general intelligence.

Chapter 2: Methodology

For the novel data collected in this thesis, the behavioural tests were chosen based on their ability to detect subtle individual differences (for example, using stimuli morphed to show facial expressions at different intensities). These stimuli had also been piloted prior to the main testing to select those that reduced the possibility of floor or ceiling effects.

For secondary data, the tests were already chosen by researchers on a separate project. This is why there are differences between papers in this thesis with regards to the tests of expression recognition, identity recognition, and intelligence.

Examples of each of the visual tasks' stimuli are shown in the respective subsections below. An outline of the independence or overlap of the studies in this thesis is included at the end of this methods chapter.

2.1 Emotion Recognition

2.1.1 Face Emotion Expression Recognition

The vast majority of emotion recognition studies have used static faces as emotion recognition stimuli. Many databases and stimulus sets have been developed over the past few decades of research, and one of the most extensively used stimulus sets is the standardised black and white prototypical facial expressions set developed by Ekman and colleagues (Ekman & Friesen, 1976). This set consists of 10 Caucasian individuals (six female and four male) portraying posed expressions of the six discrete 'basic' emotions: anger, disgust, fear, happiness, sadness, and surprise. These are considered basic emotions in the sense that they are innate, recognised and expressed universally across

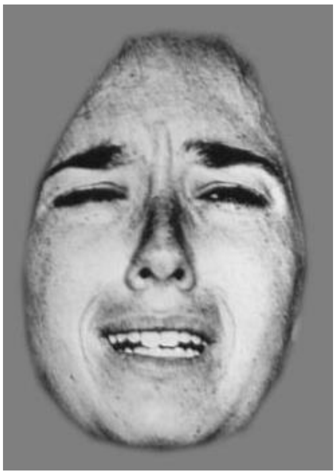


Fig 1. Exemplar image of an Ekman face (Young et al., 2002).

different cultures, and are thought to have evolved as an adaptive function to urgent and potentially life-threatening obstacles in the environment (Ekman, 1992; Izard, 1992). Given that the standardisation of this stimulus set, and its speed and ease of administering to participants, it has been widely used across emotion research since its development.

However, because these stimuli display full blown emotional expressions, ceiling effects are often observed in samples of neurotypical individuals with no emotion perception impairments. Because of this, then, some studies have morphed the facial expressions to differing intensities to increase task difficulty and individual variation (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). Another task using the Ekman faces, the Emotion Hexagon, comprises stimuli made from two discrete facial expression categories that have been morphed together to create a blended expression on a continuum between the two emotions, e.g. 90% anger, 10% disgust. The emotional categories that were blended together were always two emotions that were most easily confused with each other, and were as follows: anger-disgust, disgust-sadness, sadness-fear, fear-surprise, surprise-happiness, and happiness-anger.

The ecological validity of these stimuli have been questioned by several researchers, who argue that prototypical, static expressions are rarely encountered in ecological contexts (Isaacowitz & Stanley, 2011). They have also been criticised as being contextually impoverished, and it is plausible that this may contribute to the deficits observed in clinical populations. In real life situations, individuals who are older or who exhibit emotional difficulties may be able to compensate by using abnormal processing strategies (Adolphs, Sears, & Piven, 2001) or by using cues from other sources, such as gesture or voice (Isaacowitz & Stanley, 2011). In the context of older individuals, Ruffman (2011) argues that static emotional images could be improved upon in the future, but are still useful for research purposes. It may be informative that older adults show age-related deficits on these stimuli relative to younger adults, given that their prototypical nature and longer presentation duration may make the emotion easier to accurately recognise than the often more subtle and rapid expressions displayed in everyday interactions.

In the current thesis, facial emotion expression recognition was measured using stimuli from the Facial Emotional Expressions: Stimuli and Tests (FEEST) set that had selected items from Ekman and Friesen's (1976) Pictures of Facial Affect series. The FEEST images had been morphed so as to systematically differ in intensity between neutral and full blown expressions, and piloted in a previous study, so that items with suitable accuracies (i.e. no floor or ceiling effects) were able to be selected (Lewis et al., 2016). This meant that task difficulty was appropriately increased and subtle individual differences within neurotypical samples were able to be observed.

2.1.2 Body Emotion Expression Recognition

Much less empirical research has focused on emotion expressions conveyed by the body, relative to face or voice expression recognition. Of those that have, a variety of approaches have been taken to display bodily expression stimuli of the basic emotions, including static images of body postures, and dynamic videos of body movements. Dynamic body expressions are argued to be more ecologically valid than static displays, and are able to convey a greater amount of information to the receiver. For this reason, empirical findings suggest that dynamic displays are more accurately and efficiently recognised than their static counterparts (Atkinson, Dittrich, Gemmell, & Young, 2004).

One type of dynamic body expression is point-light displays (Brownlow, Dixon, Egbert, & Radcliffe, 1997). This type of display was first devised by Johansson (1973), and is generated by placing a number of light patches on joints of the body. The static form information is then removed, leaving only illuminated dots representing the location of the joint

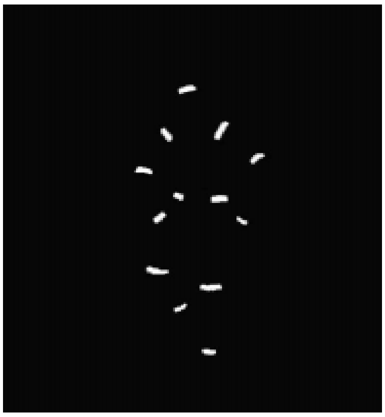


Fig 2. Exemplar image of a point light display (Atkinson et al., 2004).

patches. When static, the configuration of dots can look somewhat random, but when moving in a short video clip, the dots reliably give the percept of a dynamic individual. These point-light displays are sufficient for individuals to perceive gender, as well as detecting a familiar person through their gait or detecting a type of action being conveyed. In emotion research, the basic emotions can be reliably recognised from point-light displays (Walk & Homan, 1984), although differences in mean accuracies between the emotions have been noted (Atkinson et al., 2004).

In this thesis, body emotion expression recognition was measured using point-light displays developed by Atkinson and colleagues (2004). These displays comprised 10 actors portraying the five basic emotions (anger, disgust, fear, happiness, sadness) at three levels of emotional intensity (typical, exaggerated, very exaggerated), and were piloted in a previous study to select items suitable for individual differences research (Lewis et al., 2016).

2.1.3 Voice Emotion Expression Recognition

A considerable amount of research exists on vocal expression recognition, and findings suggest that emotional categories can be accurately recognised from the voice (Schröder, 2003). An effective way of measuring vocal expression recognition of the basic emotions is through nonverbal bursts of affect, given that these require minimal linguistic processing and can be used across different cultures to test the universality of emotion hypothesis in the auditory domain.

In this thesis, vocal emotion expression recognition was measured using stimuli drawn from the Montreal Affective Voices (MAV) set (Belin, Fillion-Bilodeau, & Gosselin, 2008). This set consists of 10 actors (five males) portraying nine emotions (anger, disgust, fear, pain, sadness, surprise, happiness, pleasure and neutral) in short nonverbal audio bursts, for example, screams or laughs. For the purposes of comparison with the face and body stimuli, the audio bursts corresponding to the five basic emotions were selected, although it should be noted that the ‘pleasure’ bursts were used as a proxy for happiness, due to pervasive ceiling effects observed in the happiness category. An additional strength of using the MAV set is its design as a counterpart to the Ekman facial expression stimuli, and therefore is a suitable measure for comparing recognition accuracy across sensory modalities (Belin et al., 2008). Given that a fundamental research question of this thesis is the multi-modal nature of emotion recognition ability, it seems pertinent to choose stimuli sets that are somewhat comparable.

2.2 Identity Recognition

Several tests of face identity recognition are presented within this thesis. Their overall format, strengths, and weaknesses are discussed below.

2.2.1 Benton Facial Recognition Test

The Benton Facial Recognition Test is a widely used test of unfamiliar face recognition, and the long form of the test involves 22 trials (and 54 responses) (Benton & Van Allen, 1968). A short form of the test comprising 13 trials (27 responses) was later developed (Levin, Hamsher, & Benton, 1975), and since then, the relative ease and brevity of administering the test has made it an established choice amongst face perception researchers.

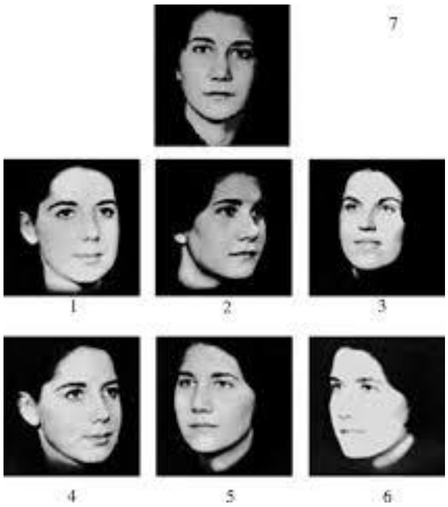


Fig 3. Exemplar image of a Benton target face [top image] and six test faces (Levin et al., 1975).

The format of the test is as follows; for every trial, the participant is shown a target face, and below this, six test faces. The participant is required to indicate which of the test faces matches the identity of the target. In the first six trials, there is only one face that matches the target identity and this image is identical to the target. In the following seven trials, three of the six test faces match the target, and there is a change in viewpoint or lighting. The faces are cropped so no clothing and little hair is shown, and no time limit is imposed upon participants who can take as long as they need to respond.

Whilst the test has been widely used, it has attracted some criticism regarding the possibility of individuals being able to use feature matching (e.g. using hairline or eyebrows) to obtain scores within 'normal' range on the test, even when such individuals have a diagnosis of prosopagnosia (Duchaine & Nakayama, 2004). Despite such individuals not appearing to use holistic face perception as in neurotypical face recognition, the simultaneous presentation of target and test faces means that they were able to match faces in a piecemeal manner (without relying on holistic face perception or face memory abilities) and obtain scores at or above the normal threshold. Duchaine and Weidenfeld (2003) thus suggested that normal Benton scores are possible without intact face recognition processes.

However, it has also been acknowledged that normal scores obtained through featural matching takes the prosopagnosic individuals a much longer time to complete the test than typical individuals, and that should a time limit be enforced, differences in normal and prosopagnosic performance may be better identified (Duchaine & Weidenfeld, 2003). Other strengths of the test have been highlighted: accuracy of healthy samples has been reported as between 81 and 83%, which suggests that whilst scores are much above chance level, they also fall below ceiling performance, and are indicative of sensitive discrimination between typical and non-typical performance (Rossion & Michel, 2018). Within the range of *typical* performance, the test has been shown to identify considerable individual variation in performance (Schretlen, Pearlson, Anthony, & Yates, 2001). Because of these reasons, the test has substantial use in both clinical and neurotypical contexts. It has been used in the current thesis in two chapters (Chapter 4 and Chapter 7) as a perceptual measure of unfamiliar face matching ability.

2.2.2. Cambridge Face Memory Test

The Cambridge Face Memory Test (CFMT) assesses memory for unfamiliar faces, and was developed as an improvement for diagnosing prosopagnosia on previous face perception tests, including the Benton Test of Facial Recognition, described above. The test's developers, sought to address the two key limitations of the Benton test (the unlimited presentation duration of stimuli and the simultaneous presentation of both target and test faces) (Duchaine & Nakayama, 2006).

To do this, they developed the CFMT, which requires encoding of six faces for a specified period of time, and then recognition of these learned faces in a total of 72 items across three distinct stages: firstly, recognition of the same faces as in the learning phase. Secondly, recognition of the same facial identities in different viewpoints and/or lighting, and thirdly, recognition of the same facial identities in different viewpoints and/or lighting with additional heavy visual noise. The test shows

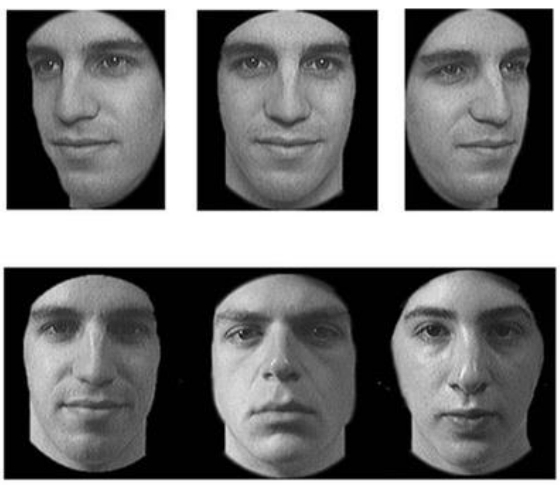


Fig 4. Exemplar image of a CFMT learning phase [upper panel] and novel testing phase [lower panel] (Duchaine & Nakayama, 2006).

absence of ceiling effects, and is able to reliably assess a wide range of individual differences in face memory ability in a neurotypical sample. Since its development, the CFMT has become one of the most widely used diagnostic assessments for prosopagnosia in clinical contexts (Bowles et al., 2009) and for assessment of individual variation in unfamiliar face memory ability in neurotypical samples. In Chapter 3, the second phase of this test is used. This was chosen as an abbreviated measure of face memory ability for purposes of keeping the survey at a reasonable length.

2.2.3 Glasgow Face Matching Test

The Glasgow Face Matching Test (GFMT) assesses unfamiliar face matching, and comprises pairs of colour face photographs, presented simultaneously in frontal view and with neutral posed expressions. The images were taken with different cameras (on the same day, approximately 15 minutes apart). Participants are shown each pair of faces one at a time and have unlimited time in which to make a same or different judgement as to the identity. The full version comprises 168 pairs of faces, whilst a shorter form of the most difficult 40 pairs is also widely used. In each version of the test, half of the face pairs are same-identity trials and half are different. The mean test accuracy was reported by the developers as 81%, although they also observed substantial individual differences in performance, ranging from 51% to 100%, making this test suitable for the research strategy of this thesis. It was chosen as a measure of unfamiliar face identity recognition in order to complement the CFMT measure of unfamiliar face learning.

2.3 General Intelligence

The very high inter-correlations of individual performance across different tests of cognitive abilities was a phenomenon first discovered over a century ago (Spearman, 1904, as cited in Duncan, Burgess, & Emslie, 2000), and was interpreted as reflecting a general factor of cognition known as Spearman's *g* (Jensen, 1998). Considerable empirical work supports the idea of *g* having a hierarchical structure of cognitive abilities (Carroll, 1993). Whilst *g* has not been without controversy in the field (Gottfredson, 1997), more recent empirical work that tested cognition in 42 tests across three different batteries found *g* factors that were completely correlated and interchangeable (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004).

This finding was replicated and extended to five independent test batteries, across which the *g* factors were very strongly associated (Johnson, te Nijenhuis, & Bouchard, 2008).

This is despite the five batteries in Johnson's 2008 paper focusing more on reasoning and manual dexterity, as opposed to the more verbal and memory components included in the batteries of Johnson et al. (2004). This lends support to *g* extending across a large range of mental ability domains.

There are a large number of measures of general intelligence available to researchers and their test of choice will depend on multiple factors, including brevity of the measure and length of their overall battery, the specific cognitive sub-area(s) of interest, the language or culture of the sample, and any other key variables they are assessing. Regardless of the chosen measure, however, research suggests that a well-established battery will reliably capture the breadth of variance associated with *g*. In this thesis, general intelligence was measured using two different tests, outlined below.

2.3.1 Alice Heim Test of General Intelligence

This test (Heim, 1970) comprises 65 items, of which approximately half are verbal in nature and half are numerical. Furthermore, half of the items require participants to type their response in an ‘open-entry’ format, whereas the other half are presented in a five-alternative multiple-choice format. Participants have a maximum time limit of 10 minutes to complete as many of the items as possible. Importantly for the research questions addressed in this thesis, it is considered a good test of *general* intelligence as it corporates elements of both crystallized knowledge (vocabulary) and fluid intelligence (online manipulation of novel information).

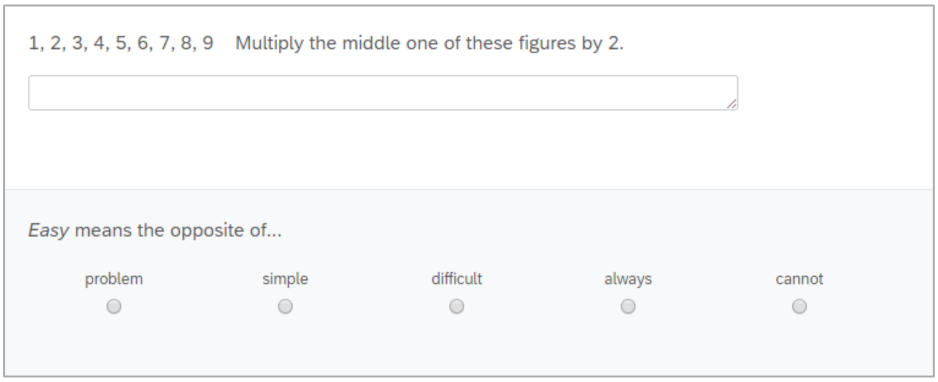


Fig 5. Exemplar item from the Alice Heim test: numerical item [upper panel] and verbal item [lower panel] (Heim, 1970).

2.3.2 Cattell Culture-Fair Intelligence Test

This nonverbal test (Cattell, 1973) comprises four sub-tests consisting of Matrices, Series Completion, Conditions, and Classification. It has traditionally been presented as a pen-and-paper test with a multiple-choice response format, and participants have set times in which to complete each subtest.

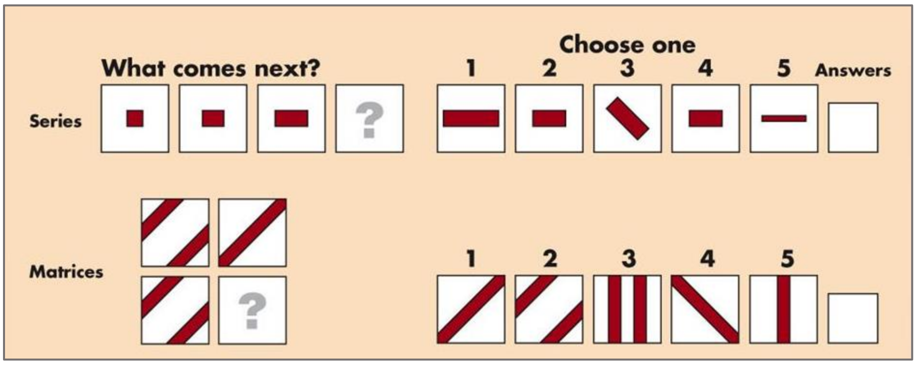


Fig 6. Exemplar items from the Cattell Culture-Fair test: Series subtest [upper panel] and Matrices subtest [lower panel] (Cattell, 1973).

Given that the Cattell Culture-Fair Test (CCFT) comprises four matrix reasoning tasks with no verbal or crystallized components, the CCFT battery has been argued by some to be solely a measure of fluid intelligence. However, whilst constructs of fluid and general intelligence have been debated in the field, factor analytic research has shown very strong correlations (.77-96) between the CCFT and other more broadly constructed cognitive batteries, e.g. the General Aptitude Test Battery (van der Gissen, 1960, as cited in Johnson et al., 2008), indicating a very high level of common measurement across these various cognitive batteries. In comparison to single measures of matrix reasoning or processing speed, then, the CCFT battery is argued to adequately capture the breadth of general intelligence.

2.4 Psychopathology

Psychopathology is measured in this thesis in healthy populations, as opposed to clinical samples, and data was gathered using self-report. Specifically, the psychopathological areas of greatest interest for this thesis comprise autism-like traits, and the sub-clinical personality construct of alexithymia.

2.4.1 Autism-like Traits

Autism was traditionally conceptualised as a clinical construct wholly distinct from the neurotypical population, but recent work points to autistic behaviours and traits being distributed on a continuum across the general population, with adults of average intelligence also reporting a range of autistic-like attributes to varying degrees of moderation (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001).

Given that a large part of the autism phenotype concerns social behaviour and interactions, it is possible that autism-like traits may contribute to emotion recognition performance, and thus this was important to control for in the current thesis. Given also that the individuals tested in this thesis are drawn from neurotypical populations, it was pertinent to choose a measure of autism-like traits suitable for use in healthy non-clinical adults.

The Autism Quotient-Short is a 28-item self-report measure (Hoekstra et al., 2011) that provides a sensitive measure of autism-like characteristics in a neurotypical population, with participants having to rate their agreement to a series of statements concerning behaviours and attributes that they may or may not experience or display. It is a widely used, freely available, and easy-to-administer measure designed for adults with average IQ or above, and has been administered to both neurotypical and clinically-diagnosed autistic populations, as well as samples of individuals with schizophrenia, prosopagnosia, anorexia, and depression (Ruzich et al., 2015).

The measure demonstrates psychometrically sound properties, with good internal consistency and test-retest reliability (Baron-Cohen et al., 2001). Importantly for testing within the general population, it has high sensitivity for discriminating between individuals low or high on autism-like traits.

2.4.2 Alexithymia

Alexithymia was also measured using a brief self-report measure: the Toronto Alexithymia Scale (TAS-20: Bagby, Parker & Taylor, 1994). This is the most widely used instrument for measuring alexithymia, with good internal consistency and test-retest reliability (Kerr, Johnson, Gans, & Krumrine, 2004). However, concerns have been raised regarding its introspective nature. Responding to the scale accurately and appropriately requires a good degree of meta-cognitive awareness regarding one’s emotional and internal state, something which is a core difficulty for alexithymic individuals (Li, Zhang, Guo, & Zhang, 2015). This raises question concerning the validity of the scale in clinical settings, particularly for those individuals at the high end of the construct. Even those with low alexithymia may rate themselves disproportionately high if they have a biased perception of others' emotional abilities (Hobson et al., 2019). Despite these concerns, the scale is the most frequently used measure in the field, and empirical data supports its validity in predicting empathic neural responses (Bird et al., 2010), emotional abilities (Lumley, Neely, & Burger, 2007), and clinical criteria.

2.5 Online surveys

Online testing has seen a surge in popularity in recent years, due to enabling data collection that is often larger (and thus better powered), less expensive, and more demographically diverse stimuli than traditional laboratory-based populations. Typical lab-based samples have tended to be homogenous in nature with low generalisability, whereas web-based participants exhibit a much greater diversity in age, sex, education, socioeconomic status, and geographic region. Web surveys also allow for collection of data from specific sub-populations, including various types of psychopathology and twin studies (Kendler, Myers, Potter, & Opalesky, 2009).

Despite some concerns over the quality and reliability of online data, results thus far appear positive. In the relevant context of face memory recognition ability, Germine and colleagues (2012) reported comparable performance on the Cambridge Memory Face Test in several large online samples and traditional lab-based samples.

Germine et al. (2012) also compared a range of other cognitive perceptual measures in web and lab-based samples, including the Reading the Mind in the Eyes test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) and Forward Digit Span from the Wechsler Adult Intelligence Scales (Wechsler, 2008). They reported comparable mean performances for the majority of the cognitive tests, and some in which the web-based sample performed more accurately, indicating no systematic differences between the online and offline samples. Thus, concerns of lack of concentration, motivation, accuracy, or honesty in unsupervised web-based participants appear to be unwarranted at least for these measures, even when the participants were anonymous and not financially compensated. This conclusion is supported by other researchers, who dispel some of the myths surrounding online data collection: Gosling and colleagues report that online responders are no less motivated, more diverse, and no more maladjusted than offline responders, and findings are in line with traditional experimental methods (Gosling, Vazire, Srivastava, & John, 2004).

2.5.1 Amazon's Mechanical Turk

One such way of collecting large-scale data online is through various crowdsourcing platforms, and the one chosen for the research conducted in this thesis was Amazon's Mechanical Turk (MTurk). This platform allows experimenters ("requesters") to advertise Human Intelligence Tasks, abbreviated to "HITs", which MTurk "workers" will select and get paid a designated sum of money ("reward") to complete. HITs are typically fairly brief and involve short surveys or marketing questionnaires, and requesters can advertise certain HITs only to workers who meet specific criteria, e.g. geographical region. Moreover, requesters can assign bonuses or reject completions, thus helping to maintain a high level of data quality (Chandler, Mueller, & Paolacci, 2014).

The demographics of MTurk workers have been found to be at least as representative of the US population as traditional samples (on measures of sex, age, race and education), and are often more diverse than the traditional lab-based undergraduate student samples (Paolacci, Chandler, & Ipeirotis, 2010). Regarding American MTurk samples, 65% of workers are reported to be female, with the average worker age being 36 years (Paolacci et al., 2010).

Given that workers earn an average hourly rate of only $1.40 an hour (Paolacci et al., 2010), the quality of collected data has been of concern to researchers. However, further research suggests this concern may be somewhat unwarranted. Firstly, workers report completing tasks for enjoyment, rather than financial reasons, and only 13% of a tested sample used MTurk as a primary source of income. Secondly, researchers have systematically compared MTurk with data from traditional US college samples and other online platforms, and found comparable results. Buhrmester, Kwang, and Gosling (2011) observed good reliabilities (mean α=.87) for MTurk data (N=3006) collected at three different levels of monetary reward, and very high test-retest reliabilities (mean r =.88) for participants tested three weeks apart, suggesting that MTurk data quality paralleled, if not exceeded, that collected in traditional samples. This finding has been supported in the literature. Casler, Bickel, and Hackett (2013) reported equivalent behavioural responses in a comparison of MTurk workers, social media responses, and traditional undergraduate lab-based sample on a learning and categorisation task, with the MTurk sample in particular being very favourably diverse. This aligns well with recent focus on moving away from traditional 'WEIRD' samples (Western, Educated, Industrialized, Rich, and Democratic) who are not representative of the population at large (Casler et al., 2013).

Another often-disputed drawback of online testing on platforms such as MTurk is the inability to monitor participant attention, concentration and motivation to complete the task seriously. However, many studies have reported successful approaches to reducing this risk in online samples, including attention check 'catch' trials, screening for series of identical responses, and as mentioned previously, examination of scale reliabilities. It has also been highlighted that the absence of a lab environment may actually be favourable in certain tasks in reducing demand characteristics, social desirability biases, and experimenter biases (Buhrmester et al., 2011). Indeed, reporting of potentially uncomfortable health, drug use, or sexual information was shown to increase with increased anonymity (Turner et al., 1998).

In a large-scale study comparing a web-based sample (N=361,703) with 510 traditional lab-based samples, Gosling and colleagues (2004) addressed six common concerns regarding Internet-collected data (demographic diversity, socially isolated samples, generalisability across presentation formats, lack of motivation, participant anonymity, and discrepancy with traditional lab-based findings). They reported that all but one (anonymity of participants) of the concerns was shown to be unwarranted. The question of anonymous responses compromising the data was considered a potentially valid concern but one that could be addressed using various approaches available to web-based platforms. In the case of MTurk, repeat responders are rare given the time and effort invested in setting up several MTurk accounts, the collection of IP addresses, and the fact that each worker ID has to correspond with a unique credit card number (Horton et al., 2010). Moreover, workers are motivated to only use one account because higher-paying HITs are restricted to workers who have a history of high-quantity, high-quality HIT completion, and so it is not in the workers' financial interest to have multiple MTurk accounts simultaneously (Chandler et al., 2014).

In sum then, Gosling and colleagues (2004) concluded that many of the earlier preconceptions regarding web-based data collection appear unfounded, and there may be substantial advantages (e.g. large, diverse samples with high internal motivation to respond appropriately) to be had over lab-based samples, that can be used in collecting demographically diverse, well-powered data in many areas of psychology.

2.6 Structural Equation Modeling

In Chapters 3 and 4 of the thesis, data is analysed using a structural equation modeling (SEM) approach. SEM is a powerful statistical technique that examines the relationships between observed (measured) and unobserved (latent) variables. It is a combination of factor analysis and multiple regression analysis and is formed of two parts: the measurement model and the structural model. The measurement model defines the relationships between both the latent constructs and their manifest indicators (measured variables) that represent the underlying theory. The structural model specifies the relationships between the latent constructs. It is considered a particularly strong technique given that it can estimate relations between latent abstract constructs in a way that is free from measurement error.

Model fit is typically assessed through a number of statistics. One of the most commonly reported is the likelihood ratio chi-square statistic, which tests whether there is a significant difference between the observed and the expected covariance matrices. Researchers typically hope that their model is not considerably different from the observed data, and thus look for a non-significant chi-square value in order to reject the null hypothesis. However, whilst the chi-square statistic is almost universally reported in studies using SEM, it has some key limitations to note. It typically can be unrealistically conservative and reject very well-fitting models if they do not meet an exact fit. Secondly, it provides a binary outcome in which the null hypothesis is either accepted or rejected. If the model is accepted, it can provide no further estimation with regards to the degree of fit (Hu & Bentler, 1998). Most problematically, it will very commonly reject models with large sample sizes, even if they provide an excellent fit to the data (Hooper, Coughlan, & Mullen, 2008). In the same vein, mis-specified models may be accepted by this statistic if they have a particularly small sample size. Based on these limitations, the chi-square continues to be reported but frequently alongside complementary fit indices (Hooper et al., 2008). Particularly with larger samples, a significant chi-square value is not necessarily interpreted as being reflective of a badly fitting model. Rather, the researcher will look to the other indices to give a more robust and indicative estimate of model fit.

As highlighted above, these other fit indices have the advantage of providing an estimate of the *degree* of model fit (or lack of) on a continuous scale, as well as being less sensitive to sample size than the chi-square value. However, they have some limitations of their own. Given the wide range of indices available, there is currently no agreed-upon standard on which of them to report, and some values can contradict each other as to the model's degree of fit, thus making it difficult for the researcher to reliably interpret and report. Hu and Bentler (1998, 1999) carried out a number of simulations of the most commonly used fit indices under differing conditions of sample size, distribution, and estimation method, and with both correct and mis-specified models. They reported that under a maximum-likelihood estimation method, a two-index strategy is recommended to be employed to reliably assess model fit. This two-step method comprises firstly the standardised root-mean-square residual (SRMR) which was the index most sensitive to simple model misspecification and was moderately sensitive to complex misspecification, although there is some indication of it being unduly affected by sample size (Cangur & Ercan, 2015). Alongside the SRMR, one of the other incremental or absolute indices should be reported, including comparative fit index (CFI) and the root-mean-square error of approximation (RMSEA). Both the CFI and RMSEA were noted to be moderately sensitive to simple misspecification and highly sensitive to complex misspecification. Recommended cut-off's for these indices are suggested as .08 for SRMR, .95 for the CFI and .06 for the RMSEA, although the authors acknowledge that concrete cut-offs are difficult to estimate given the differences in sample sizes or distributions (Hu & Bentler, 1998, 1999).

One of the more recent issues in fit indices evaluation is whether the cut-off criteria specified above and by different researchers can be generalised to different types of models. That is, two different models may be mis-specified to the same extent, and it is crucial to establish whether the fit indices would give the same model fit outcomes across these instances. Results suggest an undesirable degree of sensitivity to model type across a variety of indices, indicating that cut-off criteria would be unduly and unequally affected by differences in model structure. This poses another problem to establishing agreed-upon cut-off values across the range of indices. Of particular note, the RMSEA showed good sensitivity for model misspecification but less sensitivity to irrelevant factors of sample size and model type, suggesting this index may be a good choice for model fit index, at least with regard to differing model types. In contrast, the SRMR did show a less than ideal sensitivity to model type (Fan & Sivo, 2007). Furthermore, a recent paper advised more caution about using the SRMR, evidencing its ‘unacceptably’ low power to reject models with mis-specified covariance structures. As an alternative, the RMSEA showed very high power in these instances (Wu, West, & Taylor, 2009).

The final index to be considered here is the Akaike Information Criterion (AIC) (Akaike, 1973). This is a comparative model with the advantage of being able to assess saturated models, i.e., models with zero degrees of freedom. It also penalises for model complexity, and rewards for parsimony. It therefore strikes a fine balance between goodness-of-fit of the model and overall model simplicity. For this index, lower values indicate a superior model fit.

What appears clear from the various studies assessing fit indices is that reporting of more than one fit index for each SEM model is paramount, in order to allow for the most valid interpretation of fit (Bentler, 2007). From reviews and simulations of various fit indices, it remains difficult as to decide which indices are most appropriate and valid to include, as it may heavily depend on the model type, estimation method, normality distribution, and sample size in question. In order to provide the most well-rounded and transparent evidence possible, the structural models in Chapters 3 and 4 of this thesis will be accompanied by up to four fit indices, namely the chi-square statistic, RMSEA, CFI, and AIC.

Outline of samples in thesis

This thesis includes a combination of novel and secondary data, and this section seeks to clarify in which chapters each samples are presented.

Chapter 3: All three studies in this chapter are novel data, collected for the purpose of this thesis. These samples do not overlap with any other sample in the thesis.

Chapter 4: This sample is secondary data and was collected as part of the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) project (Shafto et al., 2014).

Chapter 5: This chapter includes a combination of novel and secondary data. The Study 1a sample came from Study 2 of Lewis et al., 2016. The Study 1b sample came from a novel collected study (no overlap with any other study in this thesis). The Study 2 sample was also a novel collected sample which was pre-registered before data collection, and which did not overlap with any other thesis study.

Chapter 6: This chapter includes a combination of novel and secondary data. The Study 1 sample comprises the combined samples from Study 1 and 2 of Lewis et al., 2016. The Study 2 sample came from Study 3 of Lewis et al., 2016. The Study 3 sample was novel collected and pre-registered before data collection, and does not overlap with any other sample.

Chapter 7: This sample was collected as part of the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) project (Shafto et al., 2014), and is the same core dataset as that used in Chapter 4. The total sample sizes differ slightly due to some participants being excluded in the Chapter 7 study because they did not have age information.

Chapter 3: Supramodal Emotion Recognition

Emotion recognition ability: evidence for a supramodal factor and its links to social cognition

The work presented in this section was conducted in collaboration with Carmen Lefevre, Andrew Young, and Gary Lewis (supervisor) and has been published in *Cognition.*

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2020). Emotion recognition ability: Evidence for a supramodal factor and its links to social cognition. *Cognition, 197*, 104166.

https://doi.org/10.1016/j.cognition.2019.104166

Abstract

Accurate recognition of others' emotions is an important skill for successful social interaction. Unsurprisingly, it has been an enduring topic of interest, and notable individual differences have been observed. Despite this focus, the underlying functional architecture of this ability has not been thoroughly investigated, particularly concerning emotion recognition across different sensory domains and stimulus modalities. Using a structural equation modelling approach, Study 1 (N=284) established the structure of emotion recognition ability across three expressive domains – face, body and voice – and observed strong evidence for a superordinate 'supramodal' emotion recognition factor, over and above domain-specific factors. Additionally, we observed a significant moderate negative association between this superordinate factor and alexithymia. In Study 2 (N=218), findings indicated that supramodal emotion recognition ability and face identity recognition are two related but independent constructs. In Study 3 (N=249), we examined links from both supramodal emotion recognition and face identity recognition to broader cognitive ability, and observed that general intelligence was a significant predictor of supramodal emotion recognition ability. In contrast, there was no association between intelligence and face identity recognition ability. Across three independent samples, then, our findings offer strong support for an emotion recognition ability factor existing across visual and auditory domains encompassing social signals conveyed by face, body and voice, and outline its associations to broader cognitive and affective traits.

3.1 Introduction

The ability to recognise emotions underpins much of human social interaction by providing important information regarding the mental states and potential intentions of others. Although the extent to which expressions of emotion involve spontaneous or more deliberately manipulative social signals is debated (Crivelli & Fridlund, 2018), accurate perception of emotional expressions is associated with better social functioning (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006) and greater relationship well-being (Carton, Kessler & Pape, 1999). Conversely, impairment in emotion recognition is associated with neuropsychiatric and psychopathological disorders including (but not limited to) depression (Surguladze et al., 2004), social anxiety (Joormann & Gotlib, 2006), and borderline personality disorder (Fenske et al., 2015).

Unsurprisingly, then, individual differences in emotion recognition ability have been a topic of broad and enduring interest. However, much of this work has been restricted to facial expressions, and this is problematic for a number of reasons. Firstly, it is clear that emotion recognition ability is not limited to expressions communicated solely by the face; affective states are expressed through additional channels, including the body and voice. Secondly, theoretical perspectives have posited that cues to emotional expression from different modalities are often closely integrated in emotion perception (Calder & Young, 2005; Schirmer & Adolphs, 2017; Young, 2018). Data from behavioural (de Gelder & Vroomen, 2000) and brain imaging studies (de Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; Hagan et al., 2009; Hagan, Woods, Johnson, Green, & Young, 2013; Peelen, Atkinson, & Vuilleumier, 2010) support this contention. A major driver of this type of organisation may be that cross-modal integration can be particularly useful in circumstances where signals from each separate channel are ambiguous and have significant temporal demands (Bruce & Young, 2012; de Gelder & Bertelson, 2003; Young & Bruce, 2011;

Young, 2018). This is often the case in everyday life, where someone's emotional reactions may change from moment to moment, or where one channel is sometimes obscured or not available (e.g. being able to see another person's face but struggling to hear them clearly). Crucially, however, it is not yet known if individual differences in emotion recognition ability reflect this theorized integrated organisation. If, as Young (2018) has argued, emotion recognition performance reflects an ability that immediately integrates relevant cues across modalities, one should expect to see the individual differences across modalities reflecting a higher-order general latent factor. Such a factor would not offer direct evidence of an integrated system *per se*, but it would provide evidence of a coherent clustering of emotion recognition abilities, in line with recent theorising by Young (2018) and others.

Of the work to address this question to date, some support for such a model has been reported. Firstly, across two studies ($N_{\text{Study 1}} = 305$, $N_{\text{Study 2}} = 295$), Schlegel, Grandjean, and Scherer (2012) assessed participants using the Multimodal Emotion Recognition Test, comprising both facial and auditory stimuli for two versions of each of five basic emotions (i.e. anger, disgust, fear, happiness, and sadness). They analysed their data using confirmatory factor analysis, and noted the presence of a general emotion recognition ability factor that acted across the modalities of face and voice. More recently, Lewis, Lefevre, and Young (2016) – also using a confirmatory factor analysis approach – reported evidence for a general ability factor acting across the modalities of face and body.

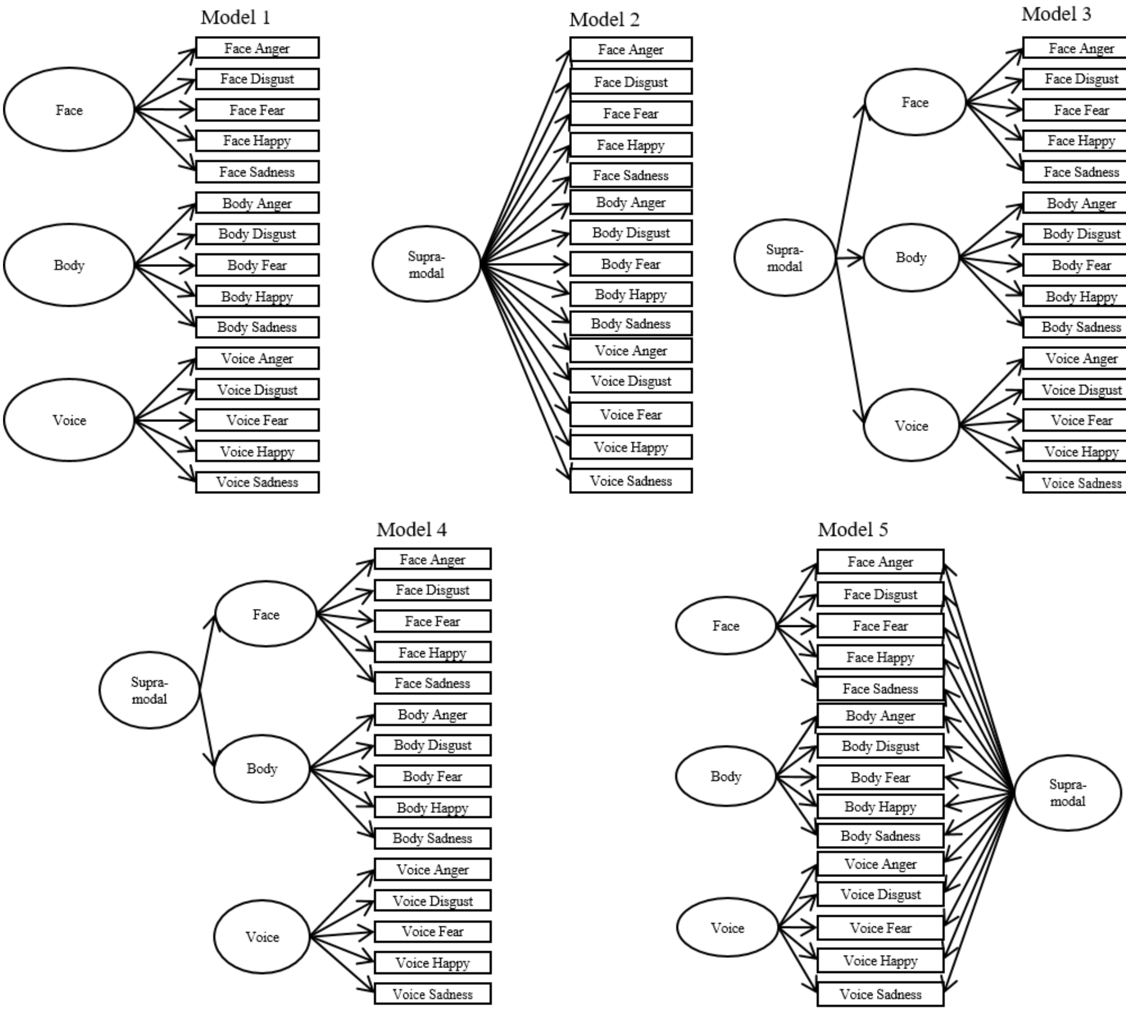
The above findings are consistent with the notion of a supramodal emotion recognition ability factor. However, a more complete test of this putative architecture requires assessment of all three main communicative channels in a single sample: the above-mentioned studies addressed only two types of stimuli (faces and voices, or faces and bodies).

More generally, if emotion recognition ability is indeed supramodal in nature, a key question is whether it reflects an emotion-specific ability or instead if it is reflective of broader socio-cognitive or general cognitive abilities. With the above in mind, in the current set of studies we sought to answer the following questions: 1) does emotion recognition ability reflect a superordinate ability factor encompassing face, body, and voice?; 2) if yes, is this supramodal ability factor specific to emotion, or reflective of broader socio-cognitive or general cognitive ability?

3.2 Study 1

In Study 1, we addressed the possible existence of a supramodal emotion recognition ability factor by examining the factor structure of emotion recognition ability across three expressive domains – face, body, and voice. We tested five models (schematically presented in Figure 1), each representing a different theoretical perspective in the field. The first model posited uncorrelated domain-specific factors underlying recognition of emotion from the face, body and voice, in line with the perspective that distinct mechanisms underlie emotion recognition from different sources of information. The second model posited a single overarching factor acting directly on all the manifest variables, in line with the idea of a general factor underlying performance across communicative modalities. The structure of the third model included latent factors for face, body, and voice emotion recognition, but also included an emotion-general factor acting directly on all three domain-specific factors (i.e. a higher-order structure). This model was posited in line with findings that emotion recognition processes are argued to be both domain-specific and domain-general across both visual (face) and auditory (voice) modalities (e.g. Lewis et al., 2016; Schlegel et al., 2012). The fourth model posited a similar higher-order structure to the third model but with the superordinate factor encompassing only the two visual domains (faces and bodies), thus allowing voice emotion recognition to be modelled as an independent factor. The final model proposed a bifactor solution whereby a supramodal factor and three domain-specific factors all contribute direct influences on the manifest variables.

Figure 1. Schematic representation of the five tested models in Study 1.



Additionally, we took the opportunity to examine whether this putative supramodal factor of emotion recognition ability was associated with a cluster of socio-affective traits with relevance for psychopathology: specifically, alexithymia, autism-like traits, and social anxiety. Our previous work (Lewis et al., 2016) highlighted a moderate negative association with alexithymia but no association with a short form measure of autism-like traits (the AQ-10). Here we sought to provide a confirmatory test of the link to alexithymia, and to further probe possible links with autism-like traits by assessing two core autism domains - Social

Behaviour/Interactions and Attention to Detail/Numbers and Patterns by using the longer 28-item Autism Quotient (AQ-28). In addition, we sought to determine whether social anxiety is linked with emotion recognition ability. This was prompted by mixed results in this domain (e.g. Mogg, Philippot, & Bradley, 2004) alongside recent work (White, Bray, & Ollendick, 2012) highlighting that social anxiety can mimic autism-like traits and thus may be the more appropriate construct to assess in the context of understanding associations between autism-like traits and emotion recognition ability.

The studies reported here were carried out on a commonly used online data collection platform: Amazon's Mechanical Turk, and which we have used extensively in our previous work (e.g. Lewis et al., 2016). Despite some researchers' early concerns about online data collection, recent studies have indicated that most of the earlier preconceptions regarding web-based data collection appear unfounded, and there may be substantial advantages to be had over lab-based samples (e.g. larger, more diverse samples with high internal motivation to respond appropriately) (Gosling, Vazire, Srivastava, & John, 2004). Specifically for Amazon's Mechanical Turk, the demographics of participants have been found to be at least as representative of the American population when compared to traditional samples (on measures of sex, age, race, and education), and are often more diverse than the traditional lab-based undergraduate student samples.

Importantly for the current paper, psychometrically sound performance in online settings has been observed in the contexts of face and voice recognition. Germine and colleagues (2012) reported comparable performance on the Cambridge Memory Face Test in several large online samples and traditional lab-based samples (Germine, Nakayama, Duchaine, Chabris, Chatterjee, & Wilmer, 2012). In a validation study of the Glasgow Voice Memory Test, Aglieri and colleagues (2017) found no significant difference between performance of an online sample (N=1120) and a controlled lab sample (N=63) (Aglieri, et al., 2017).

3.2.1 Methods

Participants

A total of 308 participants (131 males) were recruited from Amazon's Mechanical Turk service. As is the norm for online stimuli presentation, a proportion of participants experienced technical failures (e.g. videos not loading); therefore we included only those individuals who completed at least 90% (≥45 of 50) of trial blocks for each emotion and modality. In addition, we excluded participants who showed chance level performance on at least two of the emotion recognition tasks, in line with our previous work (e.g. Connolly, Young, & Lewis, 2019). This resulted in the omission of 24 participants, producing a final sample size of 284 (117 males).

Mean age of participants was 38.1 years, and ethnicity was reported as follows: White (n=222), Hispanic (n=13), Asian (n=13), Black (n=27), Native American (n=1), Arab American (n=1), Indian American (n=3) and undisclosed (n=4). These represent a typical demographic pattern for MTurk samples (Huff & Tingley, 2015). Participants gave informed consent and ethical approval was granted by the Royal Holloway, University of London Ethics Committee.

Emotion Recognition Measures

Face Emotion Recognition: To assess face emotion recognition ability, we used static images taken from the Facial Expressions of Emotion: Stimuli and Tests (FEEST) set (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). They comprise 10 identities each depicting five basic emotions (anger, disgust, fear, happiness, and sadness) morphed to different levels of intensity (i.e., N=50 images). The prototype images from which morphs were created came from the 'Pictures of facial affect' series (Ekman & Friesen, 1976). Intensity of expression was varied from neutral in 25% steps up to 125% relative to the 100% prototypical expression from the Ekman and Friesen series, using the image manipulation software Psychomorph (Tiddeman, Burt, & Perrett, 2001). These morphed stimuli were piloted in a previous study (Lewis et al., 2016), and images at varying degrees of morphed intensity were selected. This is an important step as it created appropriate levels of recognition for assessment of individual differences and in order to avoid floor/ceiling effects. This selection approach was also applied to the remaining body and voice stimulus set. The emotion recognition stimuli and task paradigm for face and body domains was the same as used in Studies 1 and 2 in Lewis et al. (2016).

Body Emotion Recognition: To assess emotion recognition ability from the body, we used point-light displays described in Atkinson, Dittrich, Gemmell, and Young (2004). This stimulus set comprises 10 different actors portraying the same five basic emotions as the face stimuli (anger, disgust, fear, happiness, and sadness), at three levels of intensity (typical, exaggerated, very exaggerated), and lasting between 4.2 and 8 seconds in length. As with the face stimuli, Lewis et al. (2016) piloted the whole range of video stimuli, and chose 50 that displayed enough variance in recognition for individual differences purposes. These stimuli were again used in our current study.

Voice Emotion Recognition: To assess vocal emotion recognition ability, we used stimuli drawn from the Montreal Affective Voices (MAV) set (Belin, Fillion-Bilodeau, & Gosselin, 2008), which comprises 10 different actors portraying nine different emotions (anger, disgust, fear, pain, sadness, surprise, happiness and pleasure, and neutral), each lasting between 1.45 and 2.23 seconds. We ensured this stimulus set was in line with the face and body sets of stimuli by including vocal bursts of only the five basic emotions (note, due to ceiling performance of the happiness vocal bursts, we instead used pleasure as a proxy for this emotion). Similarly to the other stimuli, the bursts were first piloted to arrive at 50 vocal stimuli

(10 of each emotion) with psychometric properties comparable to the other two expressive domains.

Each of these emotional modalities (face, body, voice) were included in the tested models as latent variables, with sum scores for each emotion as manifest variables.

Additional Measures

Participants were also asked to complete a series of brief questionnaires assessing the following constructs. Scale sum scores were included as manifest variables, in line with other comparable work of this kind.

Autism Quotient – Short (AQ-Short): The AQ-Short is a 28-item instrument that assesses five domains of autism-like traits: 'Social Skills', 'Routine', 'Switching', 'Imagination', and 'Attention to Detail/Numbers and Patterns'. The majority of confirmatory factor analysis studies in the field suggest that these domains can be reliably summarized by two higher order factors: i. Social Behaviour/Interactions; and ii. Attention to Detail/Numbers and Patterns (Hoekstra et al., 2011). The Social Behaviour factor (AQ-Social) reflects the first four domains and comprises 23 items: Social Skills (7 items), Routine (4 items), Switching (4 items) and Imagination (8 items). The Attention to Detail factor (AQ-Detail) reflects the final domain and comprises 5 items. Responses were made on a 4-point Likert scale, with 1 being 'definitely disagree', and 4 being 'definitely agree', with 14 of the 28 items being reverse-scored. Example items from the AQ-Social factor include "Reading a story, I find it difficult to work out the character's intentions", and examples from the AQ-Detail factor include "I am fascinated by numbers". Scale scores were generated by summing the responses from the respective sub-scales: a higher score indicates a greater degree of autism-like traits. Cronbach's alpha for our data was good for both sub-scales (Social Behaviour α=.81; Attention to Detail α=.78).

Toronto Alexithymia Scale (TAS-20): The TAS-20 is a 20-item instrument that assesses alexithymic traits in three broad sub-domains: Difficulty Identifying Feelings, Difficulty Describing Feelings, and Externally Oriented Thinking (Bagby, Parker, & Taylor, 1994). Examples of items from each subscale respectively include “I am often puzzled by sensations in my body”, “People tell me to describe my feelings more”, and “I prefer to analyse problems rather than just describe them”. Responses were made on a 5-point Likert scale, with 1 being ‘strongly disagree’, and 5 being ‘strongly agree’, with five of the 20 items being reverse-scored. Total score was generated by summing the responses from all 20 items: a higher score indicates a greater degree of alexithymic traits. Scale scores for each individual sub-scale were also generated. Cronbach’s alpha for our data was acceptable-to-good for all sub-scales (Difficulty Identifying Feelings α=.89; Difficulty Describing Feelings α=.78; Externally-Oriented Thinking α=.60).

Mini-Social Phobia Inventory (Mini-SPIN): The Mini-SPIN is a three-item measure of social anxiety disorder (Connor, Kobak, Churchill, Katzelnick, & Davidson, 2001). An example of an item is “Fear of embarrassment causes me to avoid doing things or speaking to people”. Responses were made on a 5-point Likert scale, with 1 being ‘not at all, and 5 being ‘extremely’. Total score was generated by summing the responses from all three items: higher scores indicate a greater degree of social anxiety. Cronbach’s alpha for our data was good (α=.86).

Procedure

Stimuli were blocked according to modality. Voice, face and body blocks were each presented to the participant once, and in the same block order. The within-block presentation order was fully randomised. In a five alternative-choice paradigm, participants had to select the emotion they perceived was being portrayed. Stimulus presentation consisted of a black screen for 500ms, a fixation cross for 750ms and a further 500ms black screen that preceded the onset of the stimulus. Face stimuli were presented for 1000ms. Body and vocal stimuli lasted for the duration of each individual video or audio clip. Participants could respond at any point following the stimulus onset. Following the three emotion recognition blocks, participants were asked to complete the self-report questionnaires. The participants were fully debriefed following completion of the survey.

Analysis

Structural equation models were fitted and assessed using maximum likelihood estimation with the lavaan package (Rosseel, 2012) running in R. We assessed absolute fit of our models using the Chi-Square value, the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA). The Chi-Square value has been the traditional measure of model fit but is particularly sensitive to large sample sizes and will often reject sensible models on this basis (Jöreskog & Sörbom, 1993). Therefore, we evaluated our proposed models with further indices. For the CFI and RMSEA, values of ≥.95 and ≤.06 respectively are viewed as indicative of good model fit (Hu & Bentler, 1999). Additionally, we also measured relative model fit using the Akaike Information Criterion (AIC: Akaike, 1973), with a lower AIC indicating better model fit.

3.2.2 Results

Descriptive statistics and correlations for study variables are shown in Table 1. Out of the 105 possible correlations, a total of 101 (81 of which in the expected positive direction) were significant at the 5% level. This pattern broadly conforms to previous work of this kind (Lewis et al., 2016).

Table 1. Descriptive statistics and zero-order correlations between emotions and across expressive domain in Study 1. Skew ranged from -1.47 to 0.55; kurtosis ranged from -0.80 to 1.81. Note. * $p \leq .05$; ** $p < .01$; Bold indicates $p < .001$; CI95% are presented in square brackets.

	Mean	SD	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4
Faces																
1.1 Anger	.50	.21														
1.2 Disgust	.59	.21	.18** [.07, .30]													
1.3 Fear	.67	.18	**.31** [.20, .41]	.20** [.09, .31]												
1.4 Happiness	.81	.16	**.27** [.16, .37]	.18** [.07, .30]	.17** [.05, .28]											
1.5 Sadness	.43	.21	**.31** [.20, .41]	.16** [.04, .27]	**.24** [.13, .35]	**.25** [.14, .36]										
Bodies																
2.1 Anger	.67	.22	**.37** [.27, .47]	.10 [-.02, .21]	**.26** [.14, .36]	**.31** [.20, .41]	**.34** [.23, .44]									
2.2. Disgust	.29	.18	.16** [.05, .27]	.14* [.02, .25]	.04 [-.08, .16]	-.00 [-.12, .11]	.07 [-.04, .19]	.09 [-.03, .20]								
2.3 Fear	.65	.20	**.27** [.16, .38]	.19** [.07, .30]	**.25** [.14, .36]	**.29** [.18, .39]	.19** [.07, .30]	**.43** [.33, .52]	.03 [-.09, .15]							
2.4 Happiness	.47	.18	.08 [-.04, .19]	.11 [-.01, .22]	.18** [.07, .29]	.15* [.03, .26]	.15* [.03, .26]	.17** [.05, .28]	.11 [-.01, .22]	**.23** [.12, .34]						
2.5 Sadness	.63	.21	**.33** [.23, .43]	.20** [.09, .31]	**.25** [.13, .35]	**.27** [.16, .38]	**.29** [.18, .39]	**.50** [.41, .58]	.12 [-.02, .21]	**.36** [.25, .46]	.19** [.07, .30]					
Voices																
3.1 Anger	.73	.22	**.32** [.21, .42]	.03 [-.09, .14]	.18** [.06, .29]	**.23** [.11, .33]	**.30** [.19, .40]	**.34** [.23, .44]	.15** [.04, .27]	**.28** [.17, .39]	.16** [.04, .27]	**.23** [.12, .34]				
3.2 Disgust	.77	.18	.14* [.02, .25]	.12* [.00, .23]	.18** [.07, .29]	.14* [.02, .25]	.15* [.04, .26]	.19** [.08, .30]	.11 [-.01, .22]	**.22** [.10, .33]	.08 [-.04, .20]	.19** [.07, .30]	.18** [.06, .29]			
3.3 Fear	.78	.18	.15* [.04, .27]	.08 [-.04, .19]	**.23** [.11, .33]	.12* [.01, .24]	**.26** [.14, .36]	**.29** [.18, .39]	.12* [.01, .23]	**.23** [.11, .33]	.17** [.05, .28]	**.21** [.10, .32]	.18** [.07, .29]	**.22** [.11, .33]		
3.4 Happiness	.82	.22	.08 [-.04, .20]	.03 [-.08, .15]	.12* [.00, .23]	.16** [.05, .28]	.17** [.05, .28]	.12* [.00, .23]	-.02 [-.13, .10]	.17** [.05, .28]	.16** [.02, .26]	.11 [-.01, .22]	**.22** [.11, .33]	-.01 [-.13, .11]	.18** [.06, .29]	
3.5 Sadness	.86	.15	.15* [.03, .26]	.02 [-.09, .14]	.04 [-.08, .16]	.15* [.04, .26]	.13* [.01, .24]	.19** [.08, .30]	.01 [-.11, .12]	.12* [.00, .23]	-.05 [-.16, .07]	**.21** [.09, .32]	.16** [.05, .27]	.09 [-.03, .21]	**.21** [.10, .32]	**.24** [.13, .35]

Confirmatory Factor Analyses

Model fit statistics are presented in Table 2. The first model, positing three uncorrelated and independent factors, showed at best a weak fit to the data (CFI: .61, RMSEA: .09, AIC: 17344.90). The single general factor model showed a good fit to the data (CFI: .94, RMSEA: .04, AIC: 17165.73). The third model, with a higher order structure, also showed a good fit to the data (CFI: .96, RMSEA: .03, AIC: 17155.94). The model with a general visual emotion recognition factor and a specific voice emotion recognition factor, showed a poor fit to the data (CFI: .80, RMSEA: .07, AIC: 17244.37). Finally, the model positing a bifactor structure was not identified and exhibited several improper parameter estimates: specifically, the Vocal Happiness variable showed substantial negative error variance (i.e. unlikely to be due to random sampling variability), and thus indicates model misspecification. Attempts to generate a proper model solution – e.g. omitting potentially offending indicators – were not successful and so we did not probe this model any further.

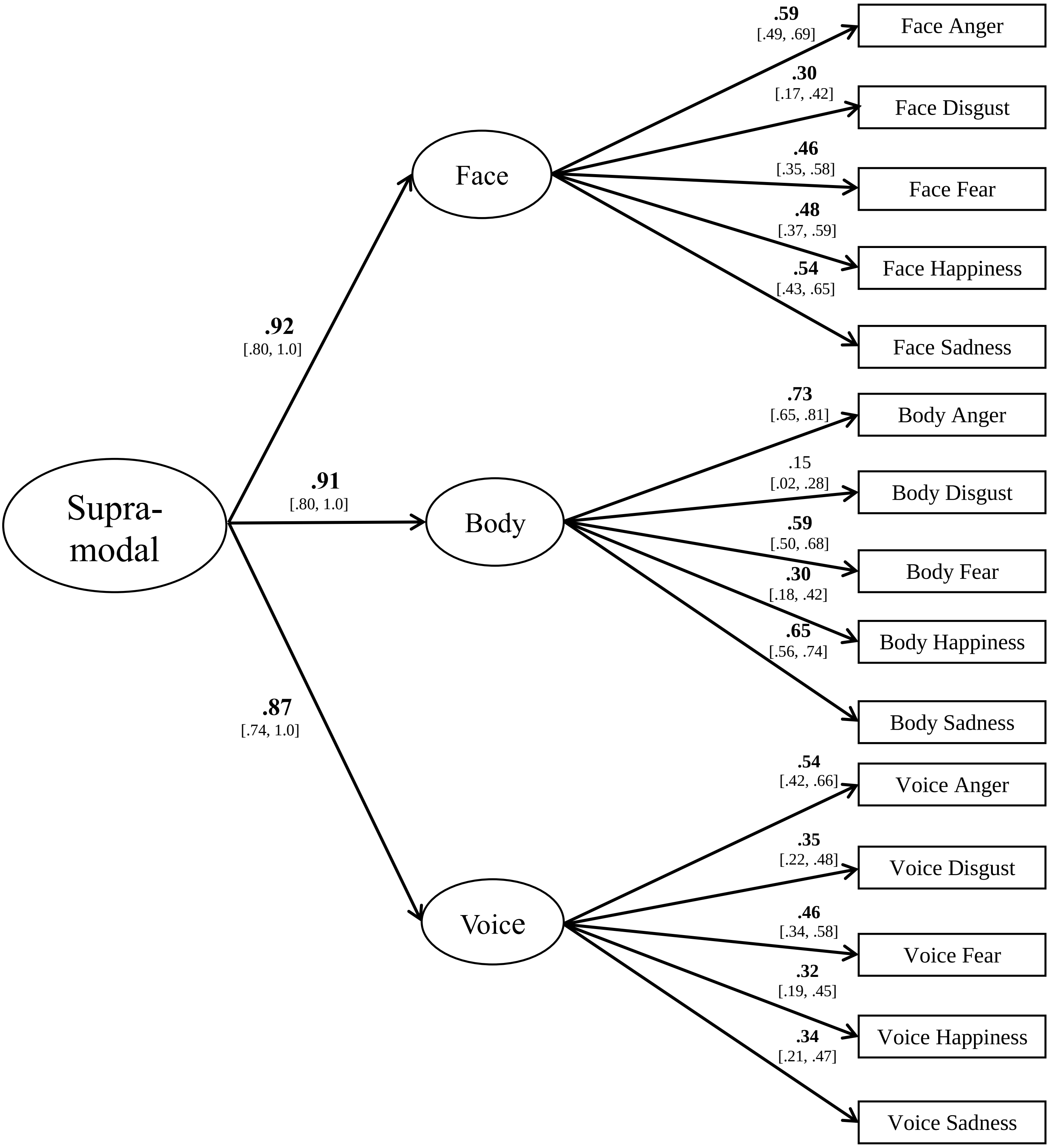
Table 2. Model output for confirmatory factor analyses in Study 1. Note that model fit values are unavailable for non-identified models, and therefore these are denoted by dashes.

Model	Description	$\chi 2$ (df)	RMSEA	CFI	AIC
1	Uncorrelated factors	304.47 (90)	.09	.61	17344.90
2	Single factor	125.31 (90)	.04	.94	17165.73
3	Higher order supramodal	109.52 (87)	.03	.96	17155.94
4	Supramodal visual modalities	199.94 (88)	.07	.80	17244.37
5	Bifactor	-	-	-	-

It is noteworthy that the model positing a single general ability factor showed only a marginally poorer fit, and the residual variance for each of the domain-specific (i.e. face, body, and voice) ability factors was modest. With this in mind, a χ^2 difference test was used as an additional point of adjudication. This analysis showed that the single factor model (with fewer estimated parameters) showed a significantly worse fit relative to the higher-order model ($\chi 2$

(3, N=284) = 15.79, *p*=.002). In sum, then, the higher-order model with a supramodal and subordinate domain-specific factors was the best fitting model and so was retained for subsequent analyses (see Figure 2).

Figure 2. Graphical representation of the final model of Study 1 parameter estimates (and 95% confidence intervals). All path coefficients in bold were significant at *p*<.02.



Relationship between Socio-Affective Traits and Emotion Recognition Ability

We next moved to considering the associations between the supramodal factor we identified and our socio-affective measures (autism-like traits, alexithymia, and social anxiety). We added in these variables individually to the higher-order model, allowing them to correlate with the supramodal factor.

We observed a moderate negative correlation between the supramodal emotion recognition factor and alexithymia (r=-.33, p<.001, [CI95%: -.41, -.25]). The associations with the three alexithymia sub-scales were as follows: Difficulty Identifying Feelings (r=-.33, p<.001, [CI95%: -.41, -.25]); Externally-Oriented Thinking (r=-.27, p<.001, [CI95%: -.34, -.20]); Difficulty Describing Feelings (r=-.19, p=.006, [CI95%: -.23, -.14]).

In contrast, no significant associations emerged for either of the Autism Quotient subscales (Social Behaviour: r=.04, p=.61, [CI95%: -.09, .17]; Attention to Detail: r=-.12, p=.08, [CI95%: -.25, .01]), or the measure as a whole (r=.00, p=.97, [CI95%: -.14, .13]). Finally, the social anxiety measure showed no association with the supramodal factor (r=-.01, p=.94, [CI95%: -.14, .12]). The correlations between the emotion recognition and affective measures are shown in Table 3.

Table 3. Associations between the emotion recognition modality blocks, and measures of autism-like traits (AQ-28), alexithymia (TAS-20) and social anxiety (Mini-SPIN) in Study 1.

	Voice		Face		Body		AQ-28		TAS-20	
	r	p	r	p	r	p	r	p	r	p
Face	**.41**	<.001								
Body	**.46**	<.001	**.53**	<.001						
AQ-28	.01	.840	.01	.887	.06	.337				
TAS-20	-.11	.056	**-.24**	<.001	**-.22**	<.001	**.49**	<.001		
Mini-SPIN	.12	.039	-.04	.559	.02	.798	**.56**	<.001	**.45**	<.001

Note. Bold indicates $p < .001$.

3.2.3 Discussion

The results of Study 1 provide evidence for a supramodal emotion recognition factor that operates across all three of the expressive domains, encompassing both visual (i.e. face and body) and auditory (voice) sources of information. This finding is consistent with recent theories that suggest emotion recognition abilities generalise across modalities (e.g. Young, 2018). It is also interesting that the results are consistent with cross-modal work (e.g. de Gelder & Vroomen, 2000) that indicates multisensory integration of complex social auditory and visual information, such as moving bodies, facial expressions, and verbal and nonverbal vocal bursts.

Additionally, we observed significant negative associations between the supramodal emotion recognition ability factor and broader socio-affective functioning: specifically, with alexithymia, as measured by the Toronto Alexithymia Scale. This observation is consistent with the findings reported by Lewis et al. (2016) and thus appears to be a robust association.

We did not find any association between the two sub-scales assessing autism-like traits (i.e. social behaviour/interactions difficulties; or attention to detail/ numbers and patterns) and supramodal emotion recognition ability. Similarly, no association was observed between social anxiety and supramodal emotion recognition ability. These results contrast with previous literature that reports biases in the recognition of specific emotional expressions (e.g. Joormann & Gotlib, 2006; Attwood et al., 2017). However, discrepancies may be due to testing different samples of participants (general or clinical population) or assessing different aspects of anxiety (state or trait). An alternative perspective comes from recent work suggesting that emotion recognition difficulties observed in autism may actually reflect alexithymia (Cook, Brewer, Shah, & Bird, 2013), which is often co-morbid with autism (Bird & Viding, 2014). The current findings are consistent with this position.

3.3 Study 2

Results from Study 1 provided evidence for the existence of a supramodal ability factor underpinning emotion recognition from the face, body, and voice. However, since this observation derives from a study involving recognition of basic emotions, it leaves open a number of questions. In Study 2, we sought to address two of these. Firstly, does this supramodal ability factor extend to naturalistic stimuli involving more complex emotional expressions? Secondly, does this supramodal factor specifically represent emotion recognition ability *per se*, or does it instead reflect broader socio-cognitive or general cognitive ability?

To examine the first question, we investigated the relationship between the supramodal ability factor and a well-used and more naturalistic test – the Reading the Mind in Films test (Golan, Baron-Cohen, Hill, & Golan, 2006) – which assesses the recognition of complex emotions and mental states from short social scenes taken from feature films.

To examine the second question, we assessed broader performance on a set of non-emotional face processing tests. For this purpose we chose the Mooney test (Mooney, 1956; Verhallen et al., 2014) as a measure of face perception, the Glasgow Face Matching Test

(Burton, White & McNeill, 2010) to measure recognition of unfamiliar face identity, and an abbreviated version of the Cambridge Face Memory Test (Duchaine & Nakayama, 2006) as a measure of face learning. These tasks thus probe different aspects of face identity perception posited in widely-used cognitive and neural models (e.g. Bruce & Young, 1986; Haxby, Hoffman & Gobbini, 2000). In order to assess whether the supramodal factor was reflective of broader cognitive ability, we also assessed participants on a short measure of verbal intelligence.

3.3.1 Methods

Participants

A total of 252 participants (104 males) were recruited from Amazon's Mechanical Turk service. As in Study 1, there were a number of participants who experienced technical failures, therefore we included only those individuals who completed 90% (≥13 of 15) of trial blocks for each emotion and modality. In addition, we again excluded participants whose responses suggested chance level responding. This resulted in an omission of 34 participants, producing a final sample size of 218 (86 males).

Mean age of participants was 37.9 years, and ethnicity was reported as follows: White (n=167), Hispanic (n=14), Asian (n=10), Black (n=16), Native American (n=3), Indian American (n=3) and Other (n=5). These characteristics match the typical demographic pattern for MTurk samples (Huff & Tingley, 2015).

Stimuli

Emotion Recognition Ability: To assess this ability, we drew on the same face, body and voice stimuli as detailed in Study 1. However, in order to reduce the time taken to complete the emotion recognition tasks and avoid any potential fatigue effects (given the additional measures used in Study 2), we selected three items from each emotion in each modality, so that each modality block comprised 15 items (5 emotions x 3 items) rather than the previous 50 (5 x 10). The selection procedure involved taking items displaying means between .50 and .80 (i.e. those items not showing floor or ceiling effects and thus suitable for individual differences research). Where a surplus of items was available, we chose those items varying in gender and age as much as possible. Where three items between this range of means were not available, we chose three that showed means closest to this range.

Reading the Mind in Films Task (RMFT): To assess the first question of Study 2, we tested participants on the more naturalistic Reading the Mind in Films test (Golan et al., 2006). This test in its full form comprises 22 short video clips taken from four British feature films, and showing scenes of characters engaging in complex social interactions. The participant perceives how a particular character is feeling from a choice of four emotional categories. Following a pilot experiment on Amazon's MTurk (n=37: 22 males), we selected only those video clips showing mean emotion recognition accuracies between .50 and .80, thus resulting in a total of 12 video clips. As before, this step is important in ensuring our stimuli sets are suitable for individual differences research.

Mooney Face Test: The Mooney test was originally developed in the late 1950's, but has recently been made suitable for online administration (Verhallen et al., 2014). The task measures perceptual closure using high contrast face images consisting of exclusively dark or light regions. Participants view an array of three images and in a three-alternative forced-choice paradigm, they are required to decide which image shows a face. In a pilot experiment involving MTurk participants (n=40: 26 males), we chose those items most suitable for individual variation (means between .70 and .88 in this case), resulting in a total of 10 items (from a total of 40 available) for the current study.

Glasgow Face Matching Test (GFMT): This unfamiliar face matching test involves showing the participant pairs of photographed faces in clear frontal view with neutral expressions and taken on the same day but photographed with different cameras. The participant has to make a same/different person judgement, with unlimited stimulus presentation time. We used the shortened version of the test comprising 40 pairs (Burton et al., 2010), and since this test has been widely used in the field (Fysh & Bindemann, 2017), we deemed it unnecessary to pilot this for use in individual differences research.

Cambridge Face Memory Test (CFMT): The original version of this unfamiliar face memory test was developed in order to diagnose individuals with prosopagnosia, and comprises a total of 72 items shown over three different phases (Duchaine & Nakayama, 2006). For purposes of keeping the total survey at a reasonable length, we chose to only use Phase 2 from this test. This phase consists of one image shown for 20 seconds at the start of the task in which six frontal-view target faces are viewed by the participant. The participant is then shown a series of 30 test images, each comprising a three-face array consisting of two distractor faces and one target face in a novel viewpoint. Their task is to select which one of the three faces they have learnt before, with unlimited time in which to make this decision.

Verbal Intelligence: The participant views a series of words (10 in total) and must choose the answer option whose meaning is closest to the word in question. The ten items in this vocabulary test were taken from the Gallup-Thorndike Verbal Intelligence Test, Form A (Thorndike, 1942, as cited in Beaujean & Sheng, 2010) and the test has been used in large scale surveys (General Social Survey: Smith, Marsden, Hout, & Kim, 2012). It offers a reasonable proxy for general intelligence since it shows a strong correlation (.87) with measures of intelligence (Jensen, 2001).

Procedure

The three emotion recognition blocks were shown to the participant first, with the same fixed order (voice, face, and body) as in Study 1. The participants then completed the face recognition tasks, in the same fixed order: GFMT, CFMT, and Mooney (each with unlimited stimulus exposure time), then the RMFT, followed by the vocabulary test. The whole procedure took approximately 30 minutes. Participants were debriefed following completion.

Analysis

Appraisal of our models was based on the same fit criteria as used in Study 1, namely the χ^2 test, the Comparative Fit Index (CFI), Root Mean Square of the Error Approximation (RMSEA), and the Akaike Information Criterion (AIC).

3.3.2 Results

Descriptive statistics and correlations between the emotion recognition variables are shown in Table 4. Out of 105 possible correlations, a total of 20 (19 of which in the expected positive direction) were significant at the 5% level. This pattern of correlations might raise concerns about the evidence for an underlying supramodal emotion recognition factor. We elected to aggregate as per our study plan, but discuss this issue in more detail below.

Table 4. Descriptive statistics and zero-order correlations between emotions and across expressive domain in Study 2. Skew ranged from -1.44 to 0.21; kurtosis ranged from -1.09 to 1.52. Note. * $p \leq .05$; ** $p < .01$; Bold indicates $p < .001$; CI95% are presented in square brackets.

	Mean	SD	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4
Faces																
1.1 Anger	.69	.28														
1.2 Disgust	.63	.27	-.01 [-.15, .12]													
1.3 Fear	.69	.29	.09 [-.04, .22]	-.02 [-.16, .11]												
1.4 Happiness	.74	.29	-.04 [-.17, .09]	.02 [-.12, .15]	.01 [-.12, .14]											
1.5 Sadness	.61	.29	.19** [.06, .32]	.04 [-.09, .18]	.01 [.15, .37]	-.03 [-.13, .14]										
Bodies																
2.1 Anger	.80	.26	.10 [-.03, .23]	.04 [-.09, .18]	.14* [.01, .27]	.13* [.00, .26]	.11 [-.03, .24]									
2.2. Disgust	.39	.26	.04 [-.09, .17]	.15* [.01, .27]	.01 [-.12, .15]	.10 [-.04, .23]	.06 [-.08, .19]	.19** [.06, .32]								
2.3 Fear	.74	.27	.06 [-.08, .19]	.00 [-.13, .14]	.11 [-.02, .24]	-.01 [-.14, .13]	-.04 [-.17, .10]	.18** [.04, .30]	-.06 [-.19, .08]							
2.4 Happiness	.67	.26	.08 [-.05, .21]	-.01 [-.14, .13]	.11 [-.03, .24]	.07 [-.07, .20]	.04 [-.09, .18]	.19** [.05, .31]	-.03 [-.16, .11]	**.28** [.13, .36]						
2.5 Sadness	.77	.25	**.25** [.13, .37]	-.03 [-.16, .11]	**.24** [.11, .36]	.00 [-.13, .14]	.10 [-.04, .23]	.11 [-.02, .24]	.13 [-.00, .26]	.08 [-.05, .21]	.02 [-.11, .15]					
Voices																
3.1 Anger	.68	.28	.08 [-.05, .21]	-.03 [-.16, .11]	-.10 [-.04, .23]	.07 [-.07, .20]	.03 [-.11, .16]	.07 [-.06, .20]	.01 [-.12, .15]	-.02 [-.16, .11]	-.08 [-.21, .05]	.19** [.06, .32]				
3.2 Disgust	.74	.24	-.02 [-.15, .11]	.02 [-.12, .15]	-.04 [-.17, .09]	.13 [-.00, .26]	.03 [-.11, .16]	.04 [-.10, .17]	.09 [-.05, .22]	-.01 [-.15, .12]	-.07 [-.20, .07]	.03 [-.11, .16]	.10 [-.03, .23]			
3.3 Fear	.82	.24	.15* [.02, .28]	-.04 [-.17, .09]	.21** [.08, .34]	.06 [-.07, .19]	.14* [.00, .26]	**.24** [.11, .36]	.06 [-.07, .19]	.06 [-.08, .19]	.15* [.02, .28]	.07 [-.07, .20]	-.10 [-.23, .04]	-.08 [-.21, .06]		
3.4 Happiness	.61	.35	.06 [-.08, .19]	-.06 [-.19, .07]	.01 [-.13, .14]	.22** [.09, .35]	-.14* [-.27, -.00]	-.02 [-.15, .11]	.09 [-.05, .22]	-.04 [-.17, .09]	.06 [-.07, .19]	-.02 [-.15, .12]	-.05 [-.18, .09]	-.03 [-.16, .10]	-.03 [-.16, .11]	
3.5 Sadness	.82	.26	.03 [-.10, .16]	.11 [-.03, .24]	.08 [-.05, .21]	-.07 [-.20, .06]	-.10 [-.23, .03]	-.01 [-.14, .13]	-.05 [-.18, .08]	-.04 [-.17, .10]	.10 [-.04, .23]	.03 [-.10, .16]	.03 [-.10, .17]	.17* [.04, .30]	.03 [-.11, .16]	.16* [.03, .29]

Correlations between the emotion recognition, face perception and verbal intelligence measures are shown in Table 5. Out of 28 possible correlations, a total of 17 were significant at the 5% level.

Table 5. Associations between the emotion recognition modality blocks, the Reading the Mind in Films Task (RMFT), the face recognition tasks (GFMT: Glasgow Face Matching Test, CFMT: Cambridge Face Memory Test, and the Mooney Face Test) and the Vocabulary test in Study 2.

	Voice		Face		Body		RMFT		GFMT		CFMT		Mooney	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
Face	**.14**	.045												
Body	.13	.054	**.29**	<.001										
RMFT	**.17**	.010	**.26**	<.001	**.26**	<.001								
GFMT	**.16**	.015	**.19**	.004	.12	.071	.09	.176						
CFMT	.10	.152	**.23**	.001	.09	.201	.13	.052	**.28**	<.001				
Mooney	-.01	.885	.06	.349	.08	.247	**.17**	.012	**.15**	.023	.08	.269		
Vocabulary	**.14**	.046	**.22**	.001	**.18**	.009	**.42**	<.001	.01	.897	**.18**	.007	**.14**	.041

Note. Bold indicates $p < .05$.

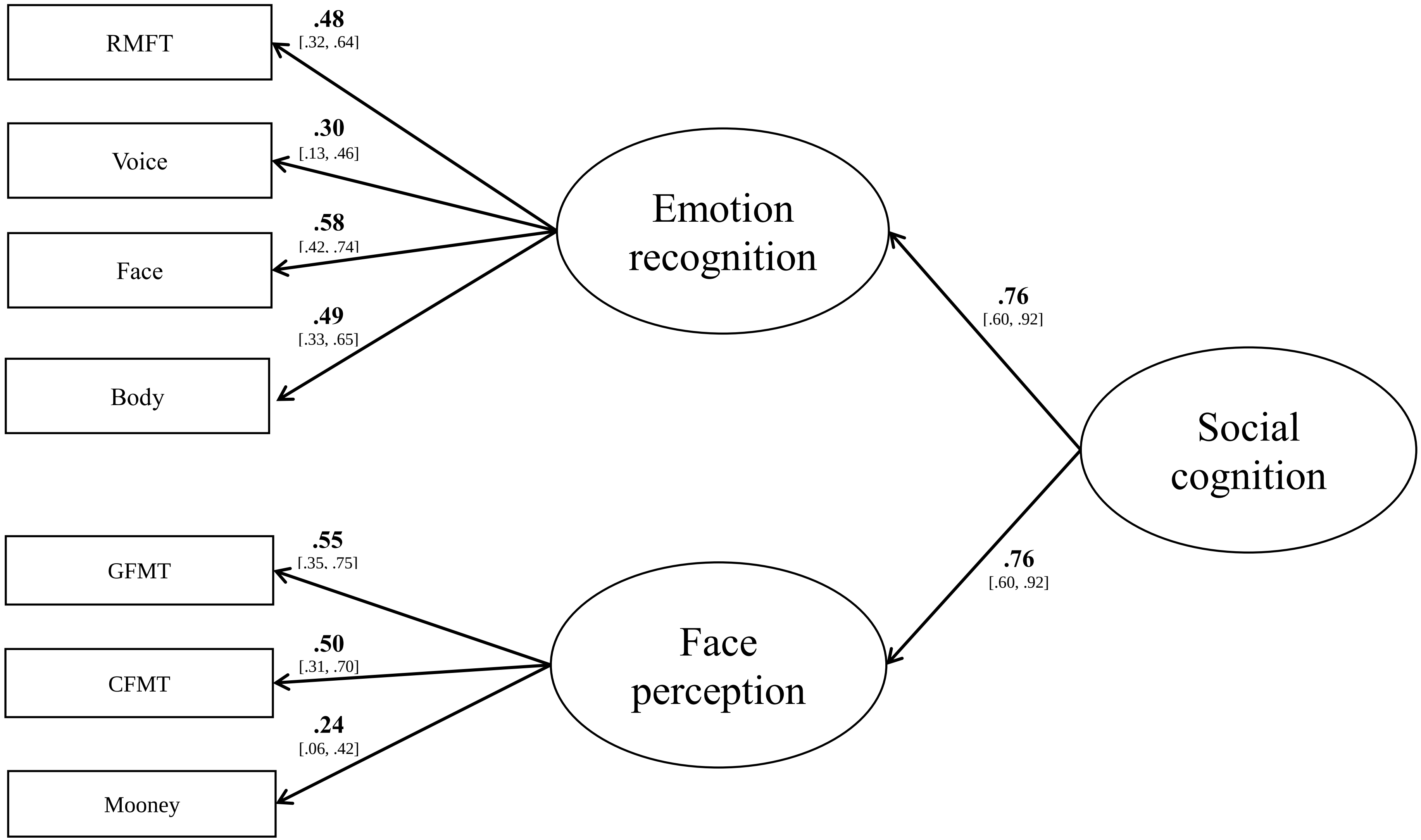
Confirmatory Factor Analyses

We first addressed whether a supramodal ability factor for emotion recognition was apparent, and particularly whether the Reading the Mind in Films Task (RMFT) showed a coherent path loading onto this general factor. A model with a latent emotion recognition ability factor loading onto face, body and voice emotion recognition and the RMFT fitted well (CFI: 1.00, RMSEA: .00, AIC: 3680.60). The RMFT showed a strong loading from the supramodal factor (.50) which was comparable to the factor loadings for the face (.53) and body emotion recognition scores (.53), and higher than the loadings of the voice scores (.28).

We next examined whether this supramodal emotion recognition factor was better understood at a more general, socio-cognitive level of abstraction, or reflected a specific ability factor. A model positing two uncorrelated latent factors – one for the emotion recognition variables and one for the non-emotion face variables – fitted the data poorly (CFI: .82, RMSEA: .07, AIC: 7233.01). A model positing a single overarching general factor fitted the data reasonably well (CFI: .92, RMSEA: .05, AIC: 7224.02). A model that specified a higher order social cognition factor loading on the emotion recognition factor and the 'non-emotional face' factor provided an excellent fit to the data (CFI: 1.00, RMSEA: .00, AIC: 7217.21). As in Study 1, a χ^2 difference test was carried out to assess whether the higher-order model showed significant fit increment relative to the single factor model. Results suggested that the single factor model showed a significantly worse fit relative to the higher-order model (χ^2 (1, N=218) = 8.82, p=.005).

Finally, we examined a bifactor solution; however, this model was not able to be identified and showed substantial negative error variance (i.e. unlikely to be due to random sampling variability). As such, we also examined a reduced bifactor model, including the general social cognition factor and either the emotion recognition factor (CFI: 1.00, RMSEA: .00, AIC: 7220.77) or face factor (CFI: 1.00, RMSEA: .00, AIC: 7220.00) – with both models exhibiting excellent fit. Overall, however, the higher order model was judged to be the best fitting model by the AIC and thus was taken forward for subsequent tests. This model, which has an overall social cognition factor and subordinate factors for emotion recognition (Face, Body, Voice, RMFT) and face perception (Mooney, GFMT, CFMT) is shown in detail in Figure 3. Of note, the emotion recognition factor forms an analogue of the supramodal emotion recognition factor observed in Study 1.

Figure 3. Graphical representation of the final model of Study 2 parameter estimates (and 95% confidence intervals). All path coefficients in bold were significant at $p<.02$. Note RMFT = Reading the Mind in Films Task; GFMT = Glasgow Face Matching Test; CFMT = Cambridge Face Memory Test.



Relationship between Verbal Intelligence and Emotion Recognition Ability

We next assessed whether verbal intelligence was predictive of the higher order general social cognition factor. A model including this parameter led to a poorly fitting model (CFI: .87, RMSEA: .06, AIC: 8062.05). We were concerned that this poor fit reflected the high verbal requirements of the RMFT, and the fact that the omission of a pathway between verbal intelligence and the RMFT essentially models these constructs as wholly unrelated; as such an additional loading directly from verbal intelligence to the RMFT was included. This model provided an excellent fit (CFI: .98, RMSEA: .02, AIC: 8048.26). Verbal intelligence was a modest-to-moderate predictor of the general factor (.40), and a modest predictor of the RMFT (.30). The final model is detailed in Figure 4, and the fit statistics for all models are shown in Table 6.

Figure 4. Graphical representation of the final model of Study 2 parameter estimates (and 95% confidence intervals) and inclusion of verbal intelligence. All path coefficients in bold were significant at *p*<.02. Note RMFT = Reading the Mind in Films Task; GFMT = Glasgow Face Matching Test; CFMT = Cambridge Face Memory Test.

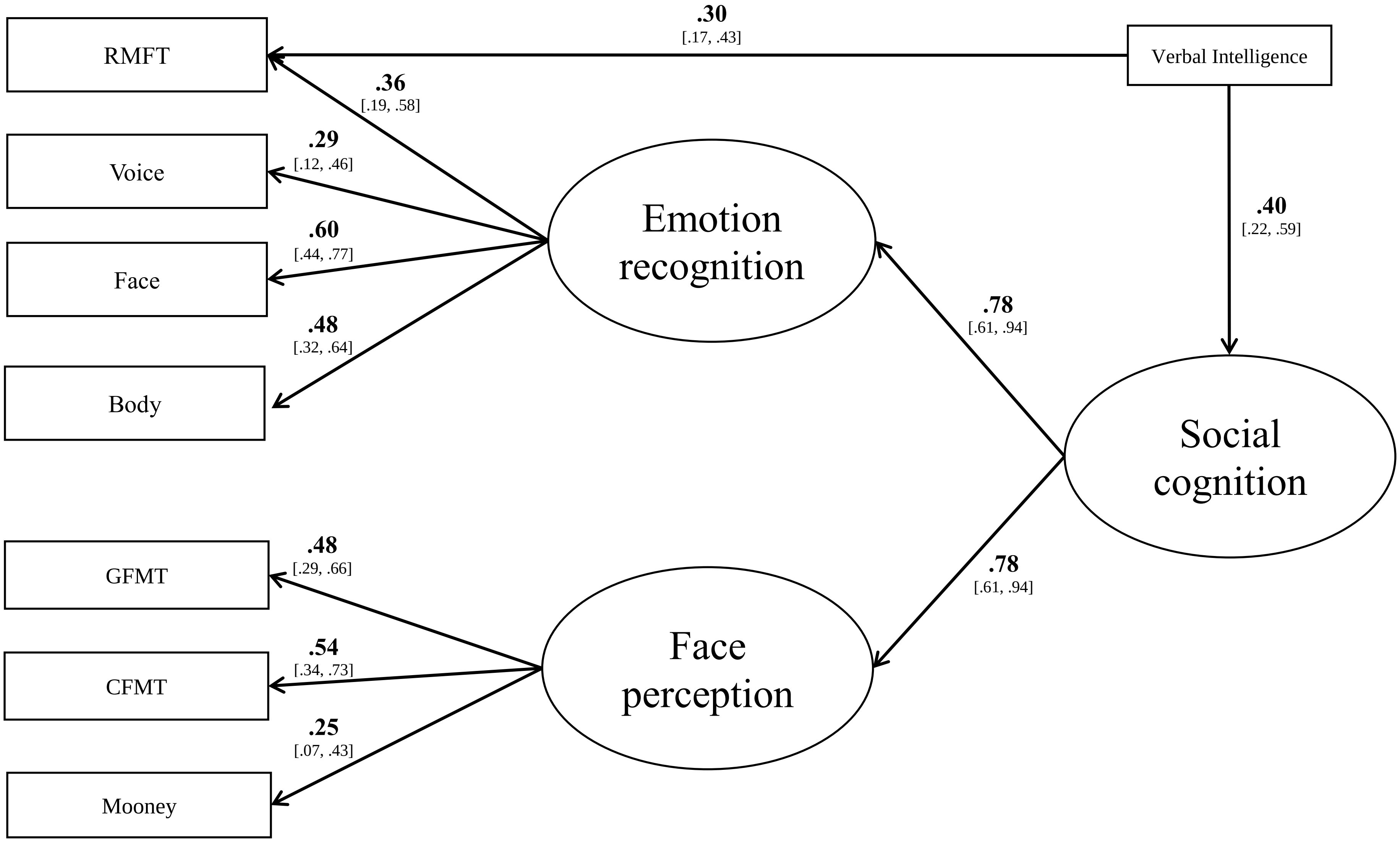


Table 6. Model output for confirmatory factor analyses in Study 2. Note that model fit values are unavailable for non-identified models, and therefore these are denoted by dashes.

Model	Description	$\chi 2$ (df)	RMSEA	CFI	AIC
1	Uncorrelated factors	29.68 (14)	.07	.82	7233.01
2	Single factor	20.70 (14)	.05	.92	7224.02
3	Higher order socio-cognitive	11.88 (13)	.00	1.00	7217.21
4	Bifactor	-	-	-	-
4a	Reduced bifactor with emotion factor	9.45 (10)	.00	1.00	7220.77
4b	Reduced bifactor with face factor	10.67 (11)	.00	1.00	7220.00
5	Higher order: verbal IQ to socio-cognitive	36.00 (19)	.06	.87	8062.05
6	Higher order: verbal IQ to socio-cognitive and RMFT	20.21 (18)	.02	.98	8048.26
7	Higher order: verbal IQ to socio-cognitive, RMFT and emotion factor	18.27 (17)	.02	.99	8048.32

3.3.3 Discussion

The results of Study 2 provide further support for the existence of a supramodal emotion recognition ability factor, and suggest that this supramodal factor extends beyond artificial laboratory-based stimuli to include more naturalistic emotional stimuli, as assessed by the Reading the Mind in Films Test (RFMT). In addition, results suggest the existence of a 'non-emotional' face perception ability factor. Most notably, our results indicate that these two latent factors represent highly related abilities – consistent with the existence of a relatively general overall socio-cognitive ability factor. Finally, we found that verbal intelligence was moderately associated with the general social cognition ability factor.

We note that the magnitude of the correlations between the individual emotion variables in Study 2 were modest and in several cases not significantly different from zero. This of course suggests that our necessarily brief measure of supramodal emotion recognition ability was noisier than desirable. Nonetheless, we still observed significant correlations across the emotion domains, indicating that these individual items did not merely capture noise, and that as well as being part of our study plan, our aggregation was reasonably principled.

3.4 Study 3

While the findings of Study 2 provide fairly clear evidence for the existence of a socio-cognitive ability factor, the brief and low-fidelity verbal intelligence measure did not allow strong assertions concerning the degree of association between socio-cognitive ability and general intelligence. With this mind, in Study 3 we sought to further probe the relationship between socio-cognitive ability and general intelligence with a broader test of intelligence that included both numerical and verbal components. Despite noting some reservations with our brief emotion recognition measure from Study 2, we elected to use the same assessment. This decision was largely a pragmatic one: including the longer emotion recognition ability measure alongside our other measures would have made our battery too long for maintaining high quality data collection. In addition, we wanted to further examine the quality of these items so as to determine if they represented a viable means to assess supramodal emotion recognition ability.

3.4.1 Methods

Participants

A total of 283 participants (147 males) were recruited from Amazon's Mechanical Turk service. There was again a proportion of participants who experienced technical failures, therefore we included only those individuals who completed 90% (≥13 of 15) of trial blocks for each emotion and modality. In addition, we excluded participants whose responses indicated chance level responding. This resulted in an omission of 34 participants, producing a final sample size of 249 (126 males).

Mean age of participants was 35.9 years, and ethnicity was reported as follows: White (n=185), Hispanic (n=12), East Asian (n=12), Black (n=26), Native American (n=2), Indian American (n=8), Middle Eastern (n=1) and Other (n=3). These participants represent the typical demographic pattern for MTurk samples (Huff & Tingley, 2015).

Stimuli and Measures

Alice Heim 4 Test of General Intelligence (AH4) Part 1 (Heim, 1970): This test consists of 65 items which are either numerical or verbal in nature. Half of the questions require multiple choice responses (from a choice of 5), and half require open entry 'creative answers'. The participants have a maximum time limit of 10 minutes to complete as many of the items as they can, after which the page automatically advances to the next section of the survey. They are encouraged to complete the questions in the order in which they are given, but are also told they may skip to the next questions if they become stuck. The participants are awarded one point for every correct answer, and their total score is generated out of a maximum of 65.

Emotion Recognition Ability: To assess this ability, we used the same abbreviated face, body and voice batteries of stimuli as detailed in Study 2. As before, each modality block comprised 15 items (3 items for each of the 5 basic emotions).

***Facial Recognition Tasks*:** The latent 'non-emotional' face factor in Study 2 showed the weakest loading onto the Mooney score, and the highest loadings onto the Glasgow Face Matching and Cambridge Face Memory tasks. Therefore we chose to use only these two latter tasks.

Glasgow Face Matching Test (GFMT): This test is fully outlined in Study 2. For time-saving purposes, here we generated an abbreviated version that was two-thirds of the length of the original Short version. Specifically, we took the 27 best performing items that showed no evidence of ceiling or floor effects and thus were suitable for individual differences research (i.e. these had means between .66 and .87).

Cambridge Face Memory Test (CFMT): This test is fully outlined in Study 2. Again, for time-saving purposes we shortened this phase to two-thirds of its original length, by including only the 20 items with accuracy means most suitable for individual differences research (i.e. these had means between .40 and .62).

Procedure

The general intelligence test was shown to the participants first. They had a maximum of 10 minutes in which to answer as many questions as possible. Participants then completed the three emotion recognition blocks, with the same fixed order (voice, face, and body) as in Studies 1 and 2. This was followed by the facial recognition tasks, in the same fixed order: GFMT and CFMT. The whole procedure took approximately 25 minutes. Participants were debriefed following completion.

Analysis

Appraisal of our models was based on the same fit criteria as used in Studies 1 and 2, namely the χ^2 test, the Comparative Fit Index (CFI), Root Mean Square of the Error Approximation (RMSEA), and the Akaike Information Criterion (AIC). Latent factors typically require three or more indicators to ensure stable identification (Floyd & Widaman, 1995). To this end, we used the first and second half of the GFMT and CFMT such that the face perception factor loaded onto four indicators. We also allowed residual correlations between the two halves of each test, as one would expect a test specific association between them. Importantly, if these associations were not formally modelled, it could lead to an unwarranted rejection of the whole model.

3.4.2 Results

Descriptive statistics and correlations between emotion variables are shown in Table 7. Out of 105 possible correlations, a total of 52 (51 of which in the expected positive direction) were significant at the 5% level. Of note, we observed a pattern of correlations between the emotion variables that more clearly supported the aggregation of these variables, although the correlations were still typically modest. This is discussed further below.

Table 7. Descriptive statistics and zero-order correlations between emotions and across expressive domain in Study 3. Skew ranged from -1.12 to 0.22; kurtosis ranged from -1.10 to 2.68. Note. * $p \leq .05$; ** $p < .01$; Bold indicates $p < .001$; CI95% are presented in square brackets

	Mean	SD	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4
Faces																
1.1 Anger	.66	.28														
1.2 Disgust	.59	.32	.14* [.03, .24]													
1.3 Fear	.67	.29	.03 [-.09, .16]	.19** [.07, .31]												
1.4 Happiness	.70	.31	.11 [-.02, .23]	.15* [.03, .27]	.08 [-.05, .20]											
1.5 Sadness	.59	.29	.19** [.07, .31]	.02 [-.10, .15]	.09 [-.03, .21]	.14* [.01, .26]										
Bodies																
2.1 Anger	.71	.31	**.31** [.20, .42]	**.27** [.15, .38]	.20** [.08, .32]	.10 [-.03, .22]	.17** [.05, .29]									
2.2. Disgust	.37	.27	.08 [-.05, .20]	.09 [-.03, .21]	.04 [-.09, .16]	.04 [-.09, .16]	.08 [-.05, .20]	**.27** [.15, .38]								
2.3 Fear	.73	.28	.14* [.02, .26]	.18** [.06, .30]	.18** [.06, .30]	.14* [.01, .26]	.21** [.09, .33]	**.34** [.23, .45]	.20** [.07, .31]							
2.4 Happiness	.65	.29	.20** [.07, .31]	.02 [-.10, .15]	.09 [-.04, .21]	.12 [-.00, .24]	.09 [-.04, .21]	**.27** [.15, .38]	**.23** [.11, .34]	**.30** [.18, .41]						
2.5 Sadness	.67	.28	**.23** [.11, .34]	.12* [.00, .25]	.13* [.01, .25]	.13* [.00, .25]	**.25** [.13, .36]	**.37** [.26, .47]	.16* [.03, .27]	**.34** [.23, .45]	.19** [.06, .30]					
Voices																
3.1 Anger	.32	.16	.08 [-.05, .20]	.06 [-.06, .18]	.10 [-.03, .22]	.05 [-.08, .17]	.12 [-.01, .24]	.08 [-.04, .21]	.07 [-.06, .19]	.06 [-.06, .19]	.02 [-.10, .15]	.12 [-.01, .24]				
3.2 Disgust	.70	.27	.05 [-.07, .18]	.10 [-.03, .22]	.06 [-.06, .19]	.07 [-.06, .19]	.07 [-.06, .19]	.18** [.06, .30]	.15* [.02, .27]	.14* [.01, .26]	.14* [.01, .26]	.07 [-.06, .19]	.02 [-.11, .14]			
3.3 Fear	.79	.27	.02 [-.11, .14]	.14* [.01, .26]	.18** [.06, .30]	.04 [-.09, .16]	.02 [-.11, .14]	**.31** [.20, .42]	.13* [.01, .25]	.19** [.07, .31]	.09 [-.03, .22]	**.26** [.14, .37]	-.13* [-.25, -.01]	.16* [.03, .28]		
3.4 Happiness	.61	.35	.07 [-.06, .19]	.19** [.07, .31]	.08 [-.04, .21]	.10 [.03, .22]	.04 [-.08, .17]	.19** [.07, .31]	.08 [-.04, .20]	.17** [.05, .29]	.04 [-.08, .17]	.14* [.01, .26]	.03 [-.10, .15]	-.02 [-.14, .11]	.20** [.01, .24]	
3.5 Sadness	.80	.27	.08 [-.04, .21]	.20** [-.08, .31]	.10 [-.03, .22]	.04 [-.08, .17]	.00 [-.12, .12]	.19** [.07, .31]	.00 [-.12, .12]	.14* [.01, .26]	.07 [-.06, .19]	.14* [.02, .26]	.10 [-.03, .22]	.10 [-.02, .22]	.11 [-.01, .23]	**.32** [.20, .42]

Correlations between the emotion recognition, face perception and general intelligence measures are shown in Table 8. Out of 15 possible correlations, a total of 11 were significant at the 5% level.

Table 8. Associations between the emotion recognition modality blocks, the face recognition tasks (GFMT: Glasgow Face Matching Test, and CFMT: Cambridge Face Memory Test, and the Alice Heim general intelligence test in Study 3.

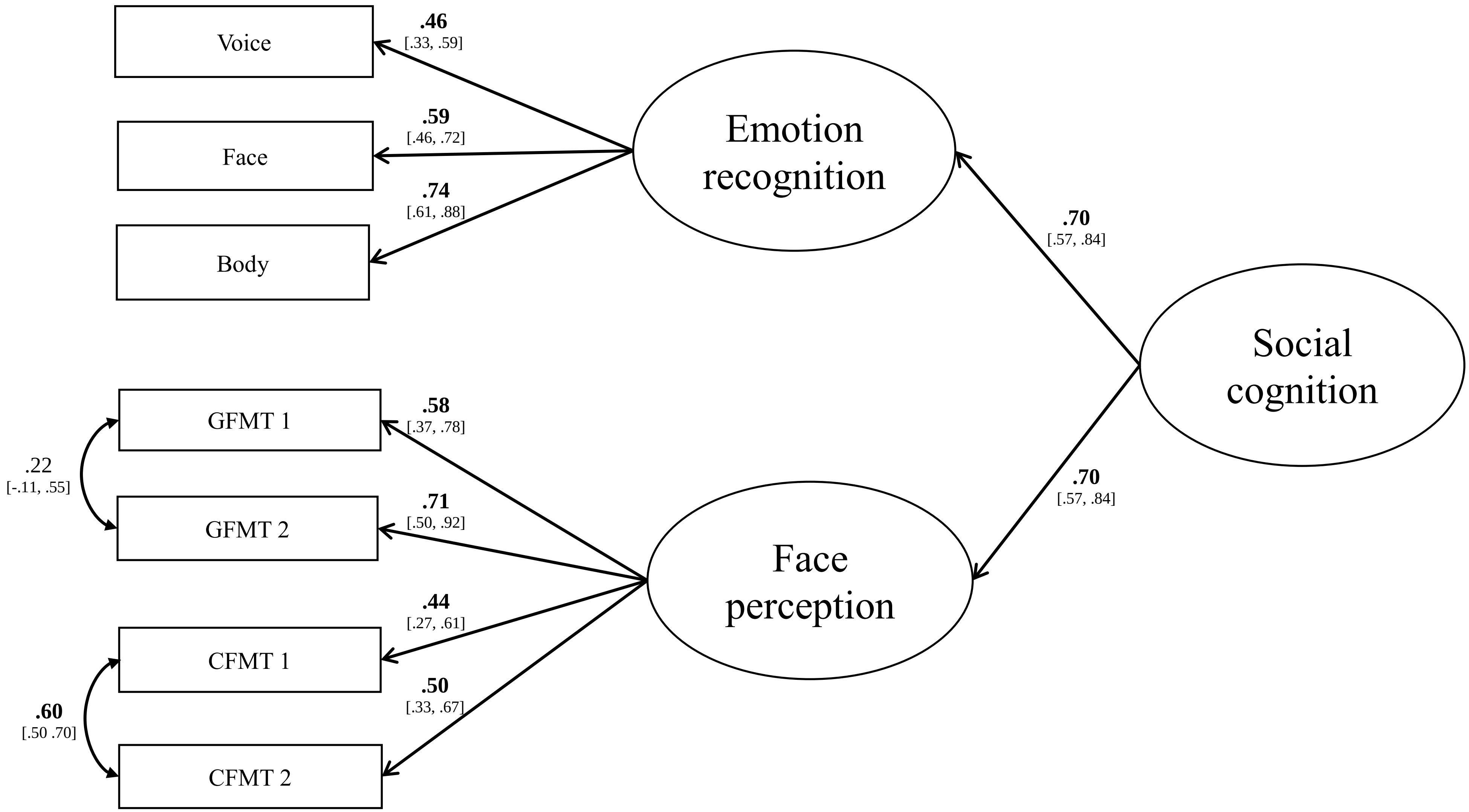
	Voice		Face		Body		GFMT		CFMT	
	r	p	r	p	r	p	r	p	r	p
Face	**.29**	<.001								
Body	**.38**	<.001	**.41**	<.001						
GFMT	-.01	.902	**.22**	.004	**.28**	<.001				
CFMT	.02	.706	**.30**	<.001	**.19**	.002	**.37**	<.001		
General Intelligence	**.38**	<.001	**.32**	<.001	**.46**	<.001	.08	.202	.10	.106

Note. Bold indicates $p < .05$.

Confirmatory Factor Analyses

We first sought to assess the fit of the models that were tested in Study 2. A model positing uncorrelated emotion recognition and 'non-emotional' face perception factors was not identified. A model positing a single overarching factor also provided a poor fit to the data (CFI: .63, RMSEA: .21, AIC: 7753.85). We next examined a model that specified a higher order social cognition factor loading on the emotion recognition factor and the face factor. This model provided a good fit to the data (CFI: .97, RMSEA: .06, AIC: 7617.56), and is shown in Figure 5. Of note, this structure mirrors that of the higher-order model in Study 2 (Figure 3), suggesting that it is robust across independent samples.

Figure 5. Graphical representation of the higher-order model of Study 3 parameter estimates (and 95% confidence intervals). All path coefficients in bold were significant at *p*<.001. Note GFMT = Glasgow Face Matching Test; CFMT = Cambridge Face Memory Test.



Finally, for purposes of completion, we assessed a bifactor model with the general social cognition factor acting directly on the manifest variables, but as in the previous two studies, this was not able to be identified. A reduced bifactor model with a general socio-cognitive factor and a specific emotion recognition factor showed an excellent fit (CFI: 1.00, RMSEA: .00, AIC: 7606.81. In contrast, a bifactor model with a general socio-cognitive factor and specific face perception factor was not identified.

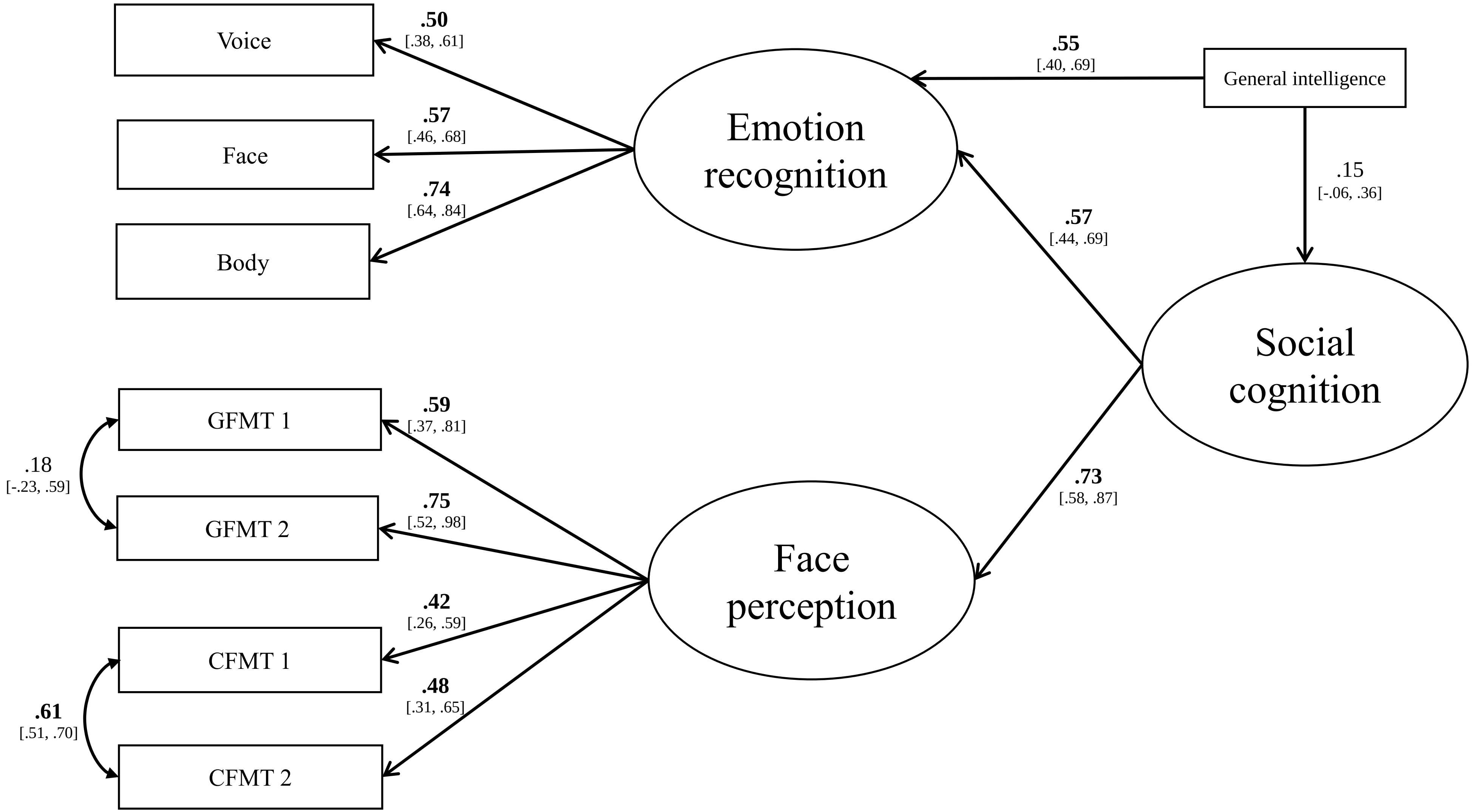
The reduced bifactor model showed a better fit compared to the higher order model, although the difference was relatively small. In line with the higher-order model showing the best fit in Study 2, we retained this model for testing the association to general intelligence, and we report the reduced bifactor model in the Supplementary Materials (Figure S1).

Relationship between General Intelligence and Emotion Recognition Ability

In the higher order model, we observed a path from general intelligence to the general socio-cognitive factor that was large in magnitude (.59); however, this model showed a relatively poor fit to the data (CFI: .91, RMSEA: .10, AIC: 9571.33). Accordingly, we explored whether general intelligence instead showed direct links to either the emotion or face factors. A model with general intelligence directly predicting the emotion recognition factor (with no path to the general socio-cognitive factor) showed a good fit to the data (CFI: .98, RMSEA: .05, AIC: 9539.48), and the pathway from general intelligence to emotion recognition was significant and substantial in magnitude (.61). In contrast, a model with general intelligence directly predicting face perception (again, with no path to the general socio-cognitive factor) did not show a good fit to the data (CFI: .82, RMSEA: .14, AIC: 9614.27), and the pathway from general intelligence to face perception was modest (-.12) and non-significant.

As an additional test, we examined a model with general intelligence directly loading on both emotion recognition and the general socio-cognitive factor. This model also fitted the data well (CFI: .98, RMSEA: .05, AIC: 9539.64). Of note, the path from general intelligence to emotion recognition was still significant and substantial in magnitude (.55), whereas the path from general intelligence to the general socio-cognitive factor was modest in magnitude (.15) and non-significant. This model is detailed in Figure 6.

Figure 6. Graphical representation of the final higher-order model of Study 3 parameter estimates (and 95% confidence intervals) and the inclusion of general intelligence. All path coefficients in bold were significant at $p<.001$. Note RMFT = Reading the Mind in Films Task; GFMT = Glasgow Face Matching Test; CFMT = Cambridge Face Memory Test.



We next performed the equivalent tests with the reduced bifactor specification. Including a direct path from general intelligence to the general socio-cognitive factor produced an unidentified model. We next examined whether a model including an additional path to the emotion recognition factor was identified. This model was identified and fitted the data well (CFI: 1.00, RMSEA: .00, AIC: 9527.17). The path from general intelligence to the emotion recognition factor was significant and substantial in magnitude (.66), whereas the path from general intelligence to the socio-cognitive factor was modest in magnitude (.12) and non-significant. This model is detailed in the Supplementary Materials (Figure S2). The fit values for all tested models are shown in Table 9.

Table 9. Model output for confirmatory factor analyses in Study 3. Note that model fit values are unavailable for non-identified models, and therefore these are denoted by dashes.

Model	Description	$\chi 2$ (df)	RMSEA	CFI	AIC
1	Uncorrelated factors	-	-	-	-
2	Single factor	163.93 (14)	.21	.63	7753.85
3	Higher order socio-cognitive	422.92 (21)	.06	.97	7617.56
4	Bifactor	-	-	-	-
4a	Reduced bifactor with emotion factor	6.89 (9)	.00	1.00	7606.81
4b	Reduced bifactor with face factor	-	-	-	-
5	Higher order: intelligence to socio-cognitive	60.66 (17)	.10	.91	9571.33
6	Higher order: intelligence to emotion factor	28.81 (17)	.05	.98	9539.48
7	Higher order: intelligence to face factor	103.60 (17)	.14	.82	9614.27
8	Higher order: intelligence to socio-cognitive and emotion factor	26.97 (16)	.05	.98	9539.64
9	Reduced bifactor: intelligence to emotion	10.49 (14)	.00	1.00	9527.17

Subsidiary Analysis of Data from Study 2

Given these findings from Study 3 of the emotion recognition factor being significantly predicted by general intelligence, we took an exploratory step of reanalysing data from Study 2 to assess whether verbal intelligence in Study 2 was more related to either the emotion recognition factor or the Reading the Mind in Films Test (RMFT) score. In a hierarchical model with verbal intelligence predicting the higher order socio-cognitive factor and the RMFT, we also included an additional path from verbal intelligence to emotion recognition, and this model showed an excellent fit to the data (CFI: .99, RMSEA: .02, AIC: 8048.32). Verbal intelligence was a significant and moderate predictor of both the socio-cognitive factor (.27) and the RMFT (.28), but showed no significant association with the emotion recognition factor. This suggests that verbal intelligence does not predict emotion recognition ability, once the association between the RMFT and broader socio-cognitive perceptual ability has been accounted for.

3.4.3 Discussion

In line with the previous two studies, the results of Study 3 offer support for the existence of a distinct supramodal emotion recognition ability factor across face, body and voice stimuli. Consistent with Study 2, a model positing two distinct, albeit highly related, factors reflecting emotion recognition and 'non-emotional' face perception abilities with a general social cognition factor in a hierarchical structure fitted the data better than a single overarching general factor model. Secondly, we noted a strong positive association between supramodal emotion recognition ability and general intelligence, broadly in line with previous studies in the field (Lewis et al., 2016; Connolly et al., 2019).

In contrast, no significant association was observed between general intelligence and either the face perception or the general socio-cognitive factors. This finding contrasts with earlier work showing an association between fluid intelligence and unfamiliar face identity matching, as measured by the Benton Facial Recognition Test (Connolly et al., 2019), but is consistent with other studies that report no correlation between intelligence and face learning, as measured by the Cambridge Face Memory Test (Wilmer et al., 2010; Palermo, O'Connor, Davis, Irons, & McKone, 2013). These studies posit that *perception* of face identity is a specific cognitive ability separate from general intelligence, and our findings here support this contention.

3.5 General Discussion

Being able to accurately interpret others' emotional expressions is an important skill for social interaction. However, to date, there has been little knowledge concerning whether individual differences in this ability reflect domain-specific or superordinate processes. In three studies, we sought to address this issue by examining the structure of individual differences in emotion recognition ability across the face, body, and voice, and the associations of this ability to various affective and cognitive traits.

The findings of these studies offer several interesting observations. Firstly, in Study 1, we provide strong support for a supramodal emotion recognition ability factor that extends beyond visual domains (facial and bodily stimuli) to also include recognition of auditory emotional stimuli from the voice. Notably, evidence for the existence of this factor was also apparent in the findings from Studies 2 and 3 and therefore appears to represent a highly replicable pattern. Secondly, we observed a moderate negative association between the supramodal emotion recognition factor and alexithymia. In contrast, no significant associations with either autism-like traits or social anxiety were seen.

In Study 2, results showed that the supramodal emotion recognition factor is also linked to recognition of complex emotional stimuli that are arguably more representative of everyday scenarios. Nonetheless, the emotion recognition factor was found to be distinct from, albeit moderately-to-strongly related to, a relatively 'non-emotional' face factor derived from various tests of face perception. Emergence of this factor is consistent with recent work reporting a general face perception factor '*f*', distinct from other cognitive abilities (Verhallen, et al., 2017; McCaffery, Robertson, Young & Burton, 2018). These authors' work assessed identity perception and other non-emotional face tasks, and therefore our observations both support the proposed existence of the *f* factor and expand its scope. Our data also suggest that *f* is related to but still somewhat distinct from other high-level visual perception abilities including emotion recognition, even when the emotional stimuli comprise facial expressions. A third latent factor that encompassed the commonality between emotion and face recognition factors in a higher order structure was posited to reflect social cognition, and this factor showed a moderate association with a brief measure of verbal intelligence.

In Study 3, results also suggested the existence of two distinct latent factors: one capturing supramodal emotion recognition ability and the other capturing face perception. As in Study 2, the commonality between these two factors was considerable and was modelled as a higher order factor that we again suggest is reflective of general social cognition. In addition, the results of Study 3 showed that general intelligence was a strong predictor of the emotion recognition factor. Accurate recognition of emotion thus appears to reflect broader, non-affective cognitive processes, and this may be for a number of reasons. In particular, our emotion recognition tasks used a multiple-choice paradigm with examples of basic emotional expressions that were selected on the basis that they were sufficiently difficult to avoid ceiling performance. As such, when participants were unsure as to the target emotion, some may have resorted to cognitive strategies to 'solve' the intended expression, such as by a process of elimination. Implementation and success of these strategies may reflect the individual's general cognitive ability. The tasks may also demand holding vocal or dynamic visual information in mind whilst simultaneously attending to five emotion labels in order to make a choice, and these working memory demands may conceivably underlie the strong relationship between emotion recognition and our test of general intelligence (Palermo et al., 2013). However, it should be noted that whilst emotion recognition and intelligence are related, they do not reflect the same construct. We note also that the presentation of the emotion stimuli lasted for the duration of each individual item (1 second for facial stimuli, between 4.2 and 8 seconds for body stimuli, and between 1.45 and 2.23 seconds for voice stimuli), and that participants had unlimited time in which to give their response. Therefore, the tests were likely capturing their ability to accurately perceive emotion, rather than simply a measure of processing speed.

In contrast to emotion recognition, no significant associations were observed between general intelligence and the general social cognition or non-emotional face perception factors. Previous research has noted the distinction between general intelligence and face perception (Wilmer et al., 2010), so the absence of association between these factors is not surprising. The face identity tests included here presented items with no time limit, and participants had as long as they needed to decide on their response. Therefore, as before, it is unlikely that processing speed or reaction time was being captured in performance on these tests.

Moreover, this further supports Verhallen et al.'s (2017) findings that neither the putative *f*' factor nor any of their individual face recognition measures show an association with scores on the standard British school qualification: *General Certificate of Secondary Education* (GCSE). In Study 3, we likewise observed no significant relationship between general intelligence and our latent face perception factor or either of the two face recognition tasks, in line with the idea that face recognition ability may be distinct from broader cognition. Our findings extend the pattern found between Verhallen's largely undergraduate-based sample and school test results to the general intelligence scores we measured in a more diverse sample of adults at the same time point as the face tasks, and suggest this independence of abilities is a robust finding and is stable over time.

Some possibilities for future research are noteworthy. Firstly, our online sample consisted solely of US participants, and so our results may not generalize to a non-Western sample. Studies of the kind reported here in broader samples would be of value. Secondly, as highlighted in Study 2, the inter-correlations of items on the shorter emotion recognition measure were limited, and as such, we advise some caution in using abbreviated scales of this kind. The limitations of abbreviated measures are a challenge for many fields of social psychology, including personality, where short form measures are commonly used, despite concerns over their validity (Bakker & Lelkes, 2018). There are clearly situations in which it is important to keep the battery length reasonable so as to avoid participant fatigue; however, the brevity of a shorter measure will often result in lower fidelity and construct validity. The balancing of test length and validity, as well as considering power, will be an ongoing challenge, and the wider community should bear this in mind when deciding which measures to include in their future studies.

Conclusion

In summary, across three independent samples, our findings provide support for a robust supramodal emotion recognition ability factor that underlies the recognition of expressions across face, body and voice, and that is in itself somewhat distinct from broader face identity perception and recognition. In addition, we have outlined the association of the supramodal emotion recognition factor to various important affective and cognitive traits, specifically alexithymia and general intelligence, as well as demonstrating that it can extend to recognising more complex and naturalistic emotional stimuli.

3.6 Supplementary Materials

Figure S1. Graphical representation of the reduced bifactor model of Study 3 parameter estimates (and 95% confidence intervals) and the inclusion of general intelligence. All path coefficients in bold were significant at *p*<.001.

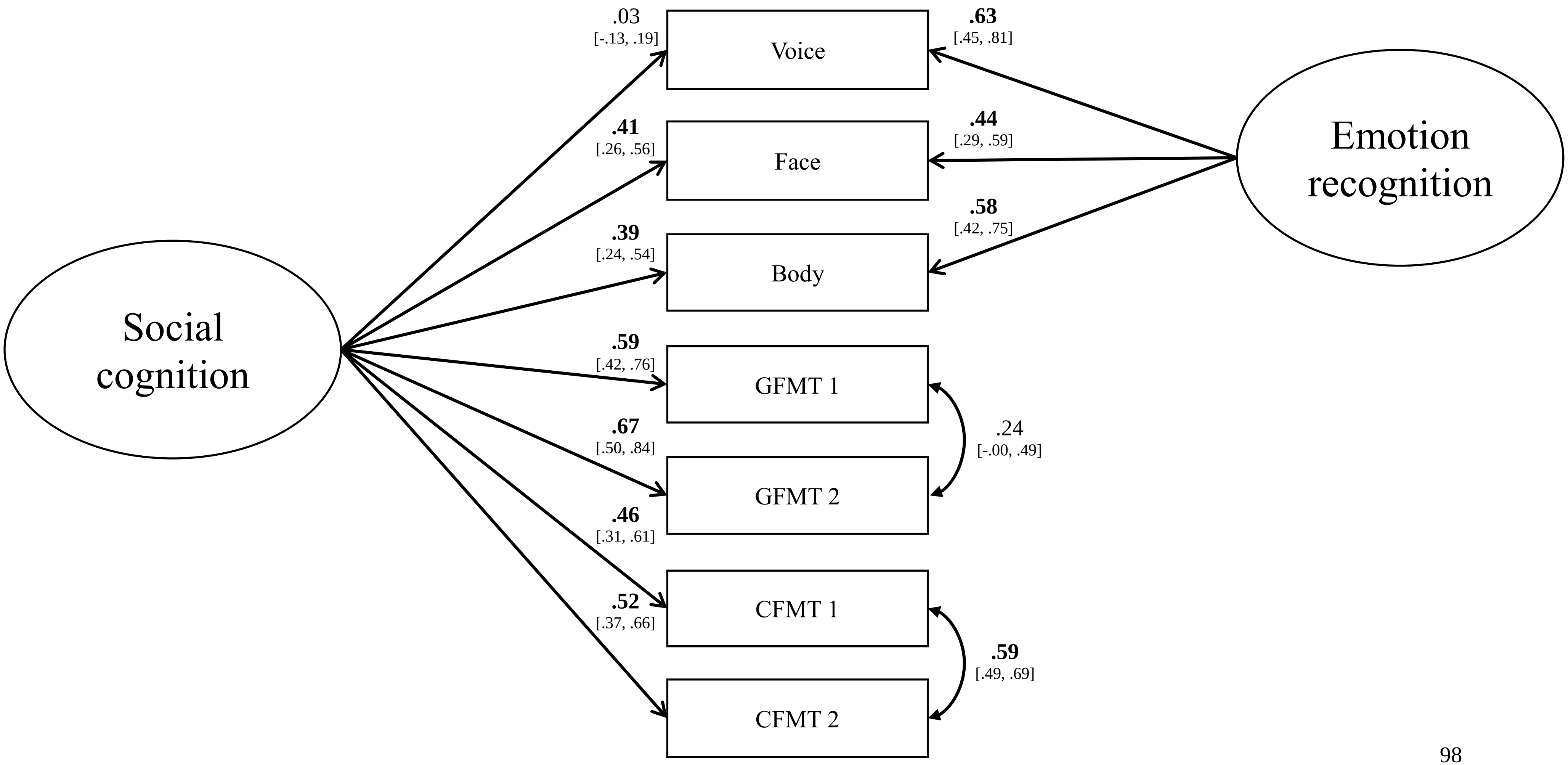
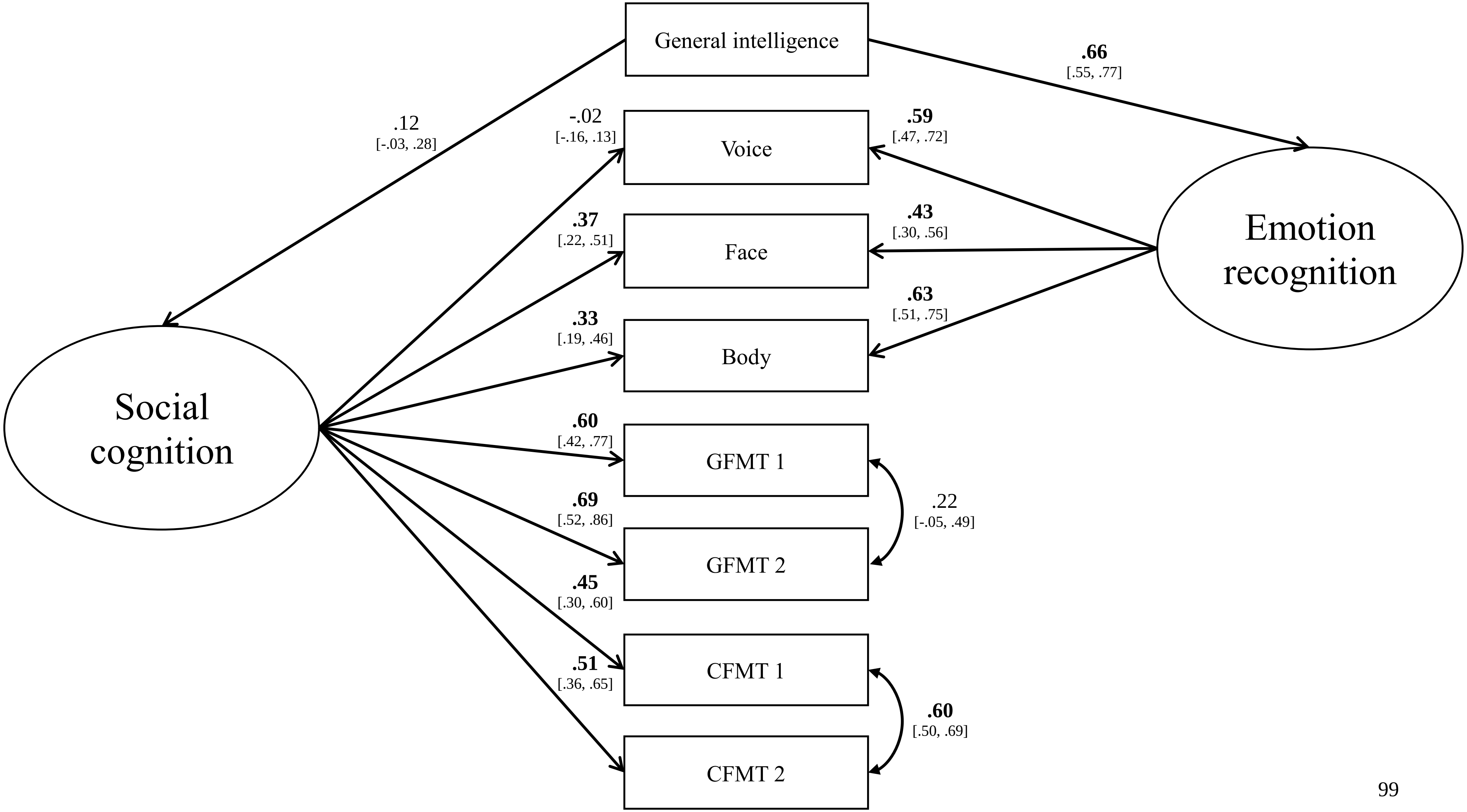


Figure S2. Graphical representation of the final reduced bifactor model of Study 3 parameter estimates (and 95% confidence intervals) and the inclusion of general intelligence. All path coefficients in bold were significant at *p*<.001.



Chapter 4: Facial Expression and Facial Identity Recognition

Recognition of facial expression and identity in part reflects a common ability, independent of general intelligence and visual short-term memory

The work presented in this section was conducted in collaboration with Andrew Young and Gary Lewis (supervisor) and is published in *Cognition and Emotion*.

Connolly, H. L., Young, A. W., & Lewis, G. J. (2019). Recognition of facial expression and identity in part reflects a common ability, independent of general intelligence and visual short-term memory. *Cognition and Emotion*, *33*(6), 1119-1128.

https://www.tandfonline.com/doi/full/10.1080/02699931.2018.1535425

Abstract

Recognising identity and emotion conveyed by the face is important for successful social interactions and has thus been the focus of considerable research. Debate has surrounded the extent to which the mechanisms underpinning face emotion and face identity recognition are distinct or share common processes. Here we use an individual differences approach to address this issue. In a well-powered (N=605) and age-diverse sample we used structural equation modelling to assess the association between face emotion recognition and face identity recognition ability. We also sought to assess whether this association (if present) reflected visual short-term memory and/or general intelligence (g). We observed a strong positive correlation ($r = .52$) between face emotion recognition ability and face identity recognition ability. This association was reduced in magnitude but still moderate in size ($r = .28$) and highly significant when adjusting for measures of g and visual short-term memory. These results indicate that face emotion and face identity recognition abilities in part share a common processing mechanism. We suggest that face processing ability involves multiple functional components and that modelling the sources of individual differences can offer an important perspective on the relationship between these components.

4.1 Introduction

Recognising someone's identity and their moods and feelings is critical to effective social interaction; identity recognition allows behaviour to be based on past experience and personal knowledge, whilst emotion recognition allows behaviour to be adjusted to current circumstances. In most everyday conditions, the face is a key source of information for both identity and emotion recognition (Bruce & Young, 2012), but the demands of these tasks differ in important ways (Young, 2018). In particular, a core requirement of face identity recognition is to be able to recognise the faces of people we know across different emotional expressions, whereas a core requirement of facial emotion recognition is to be able to recognise expressions across different identities. These differing demands imply that at some level there must be some degree of separation between mechanisms involved in recognising identity and emotion, but considerable debate has centred on the extent and nature of this functional segregation (Calder & Young, 2005).

Association between face emotion recognition and face identity recognition abilities

Classic cognitive (Bruce & Young, 1986) and neural (Haxby, Hoffman, & Gobbini, 2000) models have posited functionally distinct pathways for processing of changeable (i.e. expression) and stable (i.e. identity) facial characteristics that diverge after a common early stage in face perception. This perspective of substantial functional segregation is supported by evidence from clinical studies of brain-injured patients, which have reported double dissociations between recognition of facial expression and facial identity (Shuttleworth, Syring, & Allen, 1982; Kurucz, Feldmar, & Werner, 1979; Parry, Young, Shona, Saul, & Moss, 1991; Young, Newcombe, de Haan, Small, & Hay, 1993).

Despite this evidence of functional separation between identity and emotion recognition, other data point to some degree of overlap. For example, Rhodes et al. (2015)

reported that facial expression and identity aftereffects are positively associated, whereas no such association was observed with gaze or tilt aftereffects. In addition, fMRI studies have also been taken to argue against a complete independence of identity and expression processing; specialised brain areas thought to be involved in processing invariant facial signals (fusiform face area) and processing changeable signals (posterior superior temporal sulcus) can show sensitivity to changes in both types of facial cues (Fox, Moon, Iaria, & Barton, 2009). Similarly, interactions between identity and expression have also been noted in an early component of ERPs from neurologically normal participants (Fisher, Towler, & Eimer, 2016).

The potential discrepancy between reports of functional overlap and functional separation might be resolved if the evidence of overlap derives from the common perceptual stage posited in classic models (Bruce & Young, 1986; Haxby et al., 2000) and the evidence of separation pertains to later stages. Time-sensitive data such as ERPs suggest that this may indeed be the case (Martens, Leuthold, & Schweinberger, 2010; Fisher et al., 2016). Moreover, analysis of the image statistics that underlie recognition of identity and emotion shows that a common initial perceptual representation may be optimal. Calder, Burton, Miller, Young, and Akamatsu (2001) investigated the image statistics that underlie representations of facial identity and emotional expression through a principal component analysis (PCA) of the shape and surface properties of images from the Ekman and Friesen (1976) series of 'Pictures of facial affect'. They found that different combinations of principal components (PCs) can be used successfully to decode emotion or identity, but with two important caveats. First, multiple PCs are always needed. Second, whilst some PCs are mainly useful for decoding emotional expression and some are mainly useful for decoding identity, other PCs are useful for both identity and emotion. In other words, emotion and identity can be represented through combinations of PCs, but not in a fully exclusive manner. Hence while there are differences between the visual properties underlying identity and emotional expression, there is also substantial covariation; the principal components coding facial identity and facial expression perception show a partial overlap, rather than absolute independence.

Although some relevant lines of research already exist, then, we reasoned that the assessment of individual differences – the focus of the current study – represents a relatively novel route through which to gain further traction on this question. Existing models (Bruce & Young, 1986; Haxby et al., 2000) were formulated in an era in which it was assumed that most neurologically normal individuals will experience little difficulty in recognising facial identity and emotion. More recently, however, the existence of substantial individual differences in ability has become apparent (Duchaine & Nakayama, 2006; Burton, White & McNeill, 2010; Lewis, Lefevre & Young, 2016; Young & Burton, 2017, 2018), allowing the pattern of these differences to be informative concerning the underlying functional architecture. For example, if face emotion recognition and face identity recognition reflect contributions from a common functional component, one would expect to see an association between these abilities. Furthermore, if this common component is perceptual in nature, it should not link strongly to more general cognitive abilities (such as general intelligence), and if it is a relatively face-specific ability then it may not link strongly to other forms of perceptual aptitude (such as visual short-term memory).

Some initial work of this kind has already indicated that face emotion recognition ability and face identity recognition ability are associated. In three samples (N=40 in each) of individuals with schizophrenia, bipolar disorder, and normal health, Addington and Addington (1998) reported positive associations (r_{range} =.35-.65) between performance on the Benton Test of Facial Recognition and a facial emotion expression labelling task. Borod et al. (2000) observed a similar positive correlation of r=.44 (N=100) between these measures in a sample of healthy adults. In a study of undergraduate students (N=80), correlations of r=.40 and r=.27 were found between performance on the Cambridge Face Memory Test (CFMT: Duchaine &

Nakayama, 2006) and tests of facial emotion-matching and emotion-labelling, respectively (Palermo, O'Connor, Davis, Irons, & McKone, 2013). In a meta-analysis of 8 studies of schizophrenic individuals, a correlation of .53 was reported between facial recognition and facial emotion processing (Ventura, Wood, Jimenez, & Hellemann, 2013). Most recently, Rhodes and colleagues (2015) reported a significant positive correlation of r=.38 (N=161) between an individual's performance on the CFMT and a six alternative forced-choice facial expression recognition task.

The studies noted above provide fairly consistent evidence that individual differences in face emotion recognition ability and face identity recognition ability reflect some degree of shared processing. However, all of those examining a typical population have been based on relatively small samples (i.e. N_{range} = 40-161). As such it is possible that the reported association between face emotion and identity recognition ability is reflective of publication bias and may be inflated in its magnitude (Ioannidis, 2008) (a point we return to shortly).

Association between face emotion recognition, face identity recognition, and other cognitive abilities

To the extent that emotion and identity recognition abilities overlap, a natural question concerns whether this shared variance is reflective of face-specific ability, or instead reflects broader cognitive ability (e.g. general intelligence). A range of studies have examined the relationship between emotion/identity recognition ability and general intelligence. In the domain of emotion recognition, Lewis et al. (2016) (N=389) reported a moderate positive association (r=.43) between a latent variable of visual (face and body) emotion recognition ability and a brief general intelligence measure. Similarly, a comparable positive association (r=.44) between multimodal emotion recognition ability (simultaneous cues of face, body, and voice) and a measure of non-verbal intelligence was observed in 128 participants by Schlegel and Scherer (2016). Borod and colleagues (2000) (N=100) observed a significant positive correlation (*r*=.25) between facial emotion recognition and intelligence. In contrast, Palermo and colleagues (2013) (N=80) observed no significant association between non-verbal intelligence and either face emotion matching (*r*=.07) or face emotion labelling (*r*=.04). Similarly, in a sample of primary school-aged children (N=968), Nowicki and Duke (1994) reported that emotion recognition scores were not significantly associated with a standardized IQ measure.

A broadly equivalent pattern of findings has been noted for identity recognition ability and general intelligence. For example, work examining face memory (alongside broader face perception variables) has reported modest-to-moderate links to general intelligence using a variety of study designs, including genetic analyses (Shakeshaft & Plomin, 2015 [*r*=.16, N=718]) and structural equation modelling, as in the current study (Wilhelm et al., 2010 [*r*=.21, N=209]). Similarly using a latent modelling approach, Gignac, Shankaralingam, Walker, and Kilpatrick (2016) (N=211) observed a moderate, positive association (*r*=.34) between face recognition ability, as measured by the CFMT, and general intelligence. Finally, Hildebrandt and colleagues (2011) (N=448) reported that general intelligence accounted for approximately half of the variance of face perception and face memory abilities. In contrast, other work has found no evidence of an association between face identity recognition and other cognitive abilities (Davis et al., 2011 [N=137, *r*=-.08]; Peterson & Miller, 2012 [N=42, *r*=.01-.21]; Palermo et al., 2013 [N=80, *r*=-.01]).

While the above findings show there is still some debate on whether these associations exist between general intelligence and face emotion/identity recognition, it is clear that any test of association between identity and emotion recognition ability requires additional examination of general intelligence in order to establish the degree to which such an association is reflective of broader cognitive ability.

The Current Study

We sought to overcome some of the limitations of previous studies in the field. To this end we used a structural equation modelling approach with data from a sample of healthy, age-diverse adults, which was well-powered (see Methods) to detect even a modest association between our face emotion and identity recognition measures. Indeed, the sample size of the current study (N=605) is larger than has previously been used in the field, and can therefore offer a more robust indication of the strength of the association between face emotion and identity recognition.

In line with previous work we predicted an association of r=.40, which represented the approximate weighted mean of previous studies in the field. We also sought to test whether this association (if present) was independent of broader perceptual or cognitive processes – specifically, visual short-term memory and general intelligence. Adjusting for broader cognitive processes is necessary for our individual differences approach so that any association between face emotion and face identity recognition ability that we present is not merely an artefact of a more general ability factor. For this question, we were exploratory with our analyses given the mixed findings reported in this literature to date (Gignac et al., 2016; Lewis et al., 2016; Palermo et al., 2013; Shakeshaft & Plomin, 2015), although we expected to see some attenuation of the association between emotion and identity recognition ability.

4.2 Methods

Participants

Data analysed in this study were collected by a different research group as part of a larger project - the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) cohort (N=2681) (Shafto et al., 2014). This cohort consists of a cross-sectional adult sample (aged 18-87 years), which was randomly selected from GP listings to be demographically representative of the UK population. The cohort completed demographic questionnaires and general cognitive and memory assessments in a home interview. Following this initial assessment, 700 individuals (50 males and 50 females for every age decile) who were MRI-suitable and showed no serious cognitive impairment were invited to complete a range of neuroimaging sessions and cognitive-behavioural tasks. A total of 656 participants were thus recruited and additional analyses of their data forms the basis for the current study.

The Cam-CAN data set includes many measures. To ensure a rigorous approach we chose the measures of interest and pre-registered the approach we adopted before analysing any of the data. Note, however, that we had access to the data prior to submitting this protocol (although no analyses were performed). This was because the data were not sufficiently documented prior to receipt to be able to formulate a precise analysis plan.

In line with our pre-registration protocol (https://osf.io/e5zp8/register/565fb3678c5e4a66b5582f67), participants were excluded if they showed chance levels of performance on two or more of the cognitive-behavioural tasks, or had not completed all of the cognitive-behavioural tests (see Measures). This required the omission of 51 participants, resulting in a final sample size of 605 (291 males). The mean age of participants was 54.0 years (SD=18.2), and ethnicity was as follows: White (N=583), Asian (N=7), Black (N=1), Mixed Race (N=8) and undisclosed (N=6). All participants were native English speakers from birth.

Measures

Face Emotion Expression Recognition Test: This test is a measure of facial emotion recognition and stimuli were drawn from the Emotion Hexagon test (Young et al., 1997, 2002). This test was created by using a model from the Ekman and Friesen (1976) ‘Pictures of facial affect’ series displaying each of six basic emotions (anger, disgust, fear, happiness, sadness, and surprise). These images were each then morphed with another basic emotion to form emotional expressions with graded levels of difficulty (expression pairs morphed together consist of happiness-surprise, surprise-fear, fear-sadness, sadness-disgust, disgust-anger, and anger-happiness). In the Cam-CAN version of the test participants were shown faces with 70 or 90% of the target emotion, with a six alternative forced-choice response involving 20 trials for each of the emotions. Stimuli were shown for 3 seconds each. A percentage accuracy score for each of the six emotions was generated for use in subsequent analyses. The six Emotion Expression Recognition sub-scores were significantly associated: r ranged from .12 to .46, and all $p < .003$.

Benton Test of Facial Recognition: Participants completed the short-form of the Benton Test of Facial Recognition (Levin, Hamsher, & Benton, 1975), which measures an individual's ability to match pictures of unfamiliar faces. The test consists of 27 trials in which the participant is shown one target face and an array of six faces. The participant has to find one or more examples of the target face in the array. Changes in head orientation or lighting can occur between the target and array faces. Each correct response receives a score of 1, and a total percentage accuracy score was generated for use in subsequent analyses.

Cattell Culture Fair Intelligence Test: Participants completed the standard form of the Cattell Culture Fair Intelligence test, Scale 2 Form A (Cattell, 1973). The test contains four nonverbal subtests: Series Completion, Classification, Matrices, and Conditions, and participants are given 3, 4, 3, and 2.5 minutes to complete each subtest respectively. The Cattell test is a pen-and-paper test where the participant chooses a response on each item from multiple response options, and records responses on an answer sheet. Correct responses are given a score of 1 and the percentage correct for each sub-test was calculated for use in subsequent analyses. The four Cattell Culture Fair Intelligence subtests were significantly associated: r ranged from .53 to .64, and all $p < .001$.

Visual short-term memory (VSTM): This task assesses participants' short-term memory for colour, and is adapted from a previous experiment (Zhang & Luck, 2008). It is tested using a continuous colour-wheel paradigm, and consists of two blocks of 112 trials. On each trial, participants see a display for 250ms containing a central fixation and one to four coloured discs in the surround. A blank screen is then displayed for 900ms. Following this, one of the disc locations is highlighted and the participant has unlimited time in which to indicate the colour of the cued disc that was previously in the highlighted location using a continuous colour response wheel. The participant's accuracy of reported disc colour (precision) was generated for each learning set size (1, 2, 3 or 4 discs). Three out of the four visual short-term memory sub-scores were significantly associated with each other (r ranged from .17 to .42, all $p < .001$), although the scores for learning set size 4 were not significantly associated to the other three learning set sizes (r ranged from .02 to .05, p ranged from .23 to .67). As such, we used learning set sizes 1, 2 and 3 to form an aggregate accuracy score. This variable was used in the reported analyses. As a sensitivity check we also ran our analyses where the VSTM accuracy score was derived just from learning set size 4 (see Supplementary Materials).

Analysis

We used a structural equation modelling approach to address our research questions. Specifically, we modelled face emotion recognition ability as a latent variable loading on the six manifest variables of anger, disgust, fear, happiness, sadness and surprise recognition scores. We did not have access to the raw items for the Benton Test of Facial Recognition and so here we used the total score as a manifest variable. The latent face emotion recognition ability factor and the Benton Test score were allowed to covary in our model to assess if an association was present (see Figure 1). To assess if the association between face emotion and identity recognition ability was independent of general intelligence and visual short-term memory in a second model (see Figure 2) we regressed these variables onto a latent factor of general intelligence (which was defined by the four Cattell sub-tests) and visual short-term memory.

We examined absolute fit of our models using the χ^2 value, the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA). The χ^2 value is the traditional measure of model fit, but is very sensitive to large sample sizes and often rejects sensible models on this basis (Jöreskog & Sörbom, 1993). For the CFI and RMSEA, values of $\geq .95$ and $\leq .06$ respectively are viewed as indicative of good model fit (Hu & Bentler, 1999).

4.3 Results

Descriptive statistics and correlations for study variables are shown in Table 1. All study variables were approximately normally distributed. In summary, of the 78 possible correlations, all were positively signed and 74 were significant at the 5% level.

Table 1. Correlations between the emotion recognition scores, the Benton Test of Facial Recognition, the Cattell Fluid Intelligence test and its four subtests, and the Visual Short-term Memory (VSTM) combined accuracy score.

	Mean	SD	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Benton	Cattell Total	Series Completion	Classification	Matrices	Conditions
Anger	83.4	19.6												
Disgust	86.2	18.3	**.43**											
Fear	75.3	21.9	**.46**	**.28**										
Happiness	97.5	5.4	**.20**	**.18**	**.19**									
Sadness	92.5	13.3	**.40**	**.23**	**.43**	**.18**								
Surprise	89.5	11.1	**.26**	.12**	**.18**	**.12**	**.31**							
Benton	85.0	8.5	**.30**	.09*	**.38**	.03	**.35**	**.27**						
Cattell Total	69.5	14.6	**.43**	**.15**	**.48**	.12**	**.42**	**.33**	**.42**					
-Series Completion	78.6	16.9	**.39**	**.17**	**.41**	.12**	**.41**	**.34**	**.41**	**.84**				
-Classification	58.1	15.2	**.30**	.06	**.34**	.07	**.29**	**.22**	**.26**	**.81**	**.54**			
-Matrices	75.7	18.2	**.40**	**.15**	**.46**	.13**	**.36**	**.27**	**.43**	**.85**	**.64**	**.58**		
-Conditions	66.2	23.0	**.32**	.10*	**.37**	.08*	**.31**	**.27**	**.29**	**.79**	**.56**	**.53**	**.54**	
VSTM	1.43	.37	**.13**	.07	**.26**	.08*	.13**	.10*	**.17**	**.30**	**.24**	**.25**	**.28**	**.22**

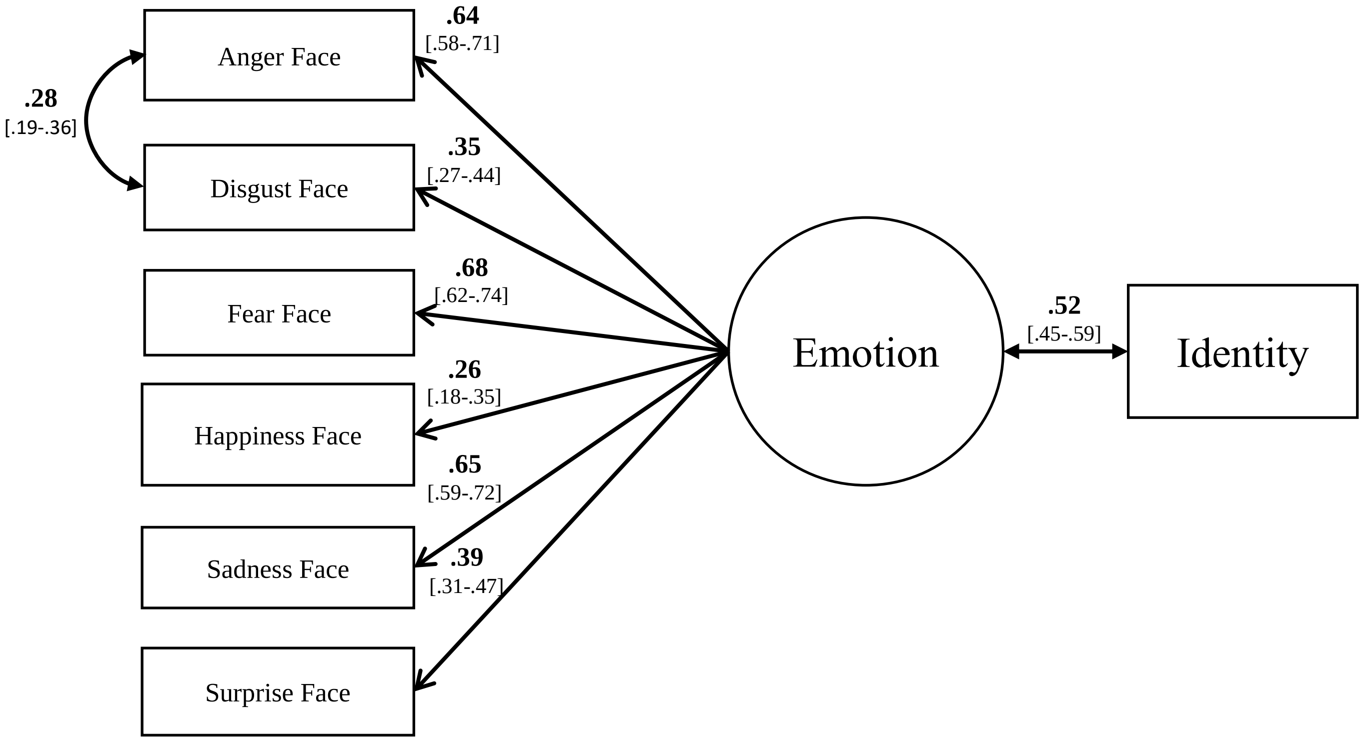
Note. Performance scores ranged from 5 to 100% accuracy. Note. * $p < .05$; ** $p < .01$; Bold indicates $p < .001$.

Confirmatory Factor Analyses

We first assessed the association between face emotion recognition ability (face emotion) and face identity recognition ability (face identity). This model showed an acceptable fit to the data (CFI: .91, RMSEA: .09), and in answer to our first research question, we observed a strong positive correlation between the face emotion and face identity recognition variables (r=.49, p<.001).

While the model fit was acceptable, it was not good by conventional standards. When global model fit is less than good, this can lead to biased estimates of local parameters (Tomarken & Waller, 2003) – in the current model we were concerned that this might inflate the estimate of the association between emotion and identity recognition. Therefore, we examined modification indices as a sensitivity analysis, in order to assess whether additional paths would improve model fit. These highlighted a covariance path between anger and disgust recognition which would enhance model fit. As such, we re-ran our model, but here including this additional parameter (see Figure 1). This model showed a good fit to the data (CFI: .95, RMSEA: .06), and we again observed a strong positive correlation between the face emotion and face identity recognition variables (r=.52, p<.001). No other model modifications were examined.

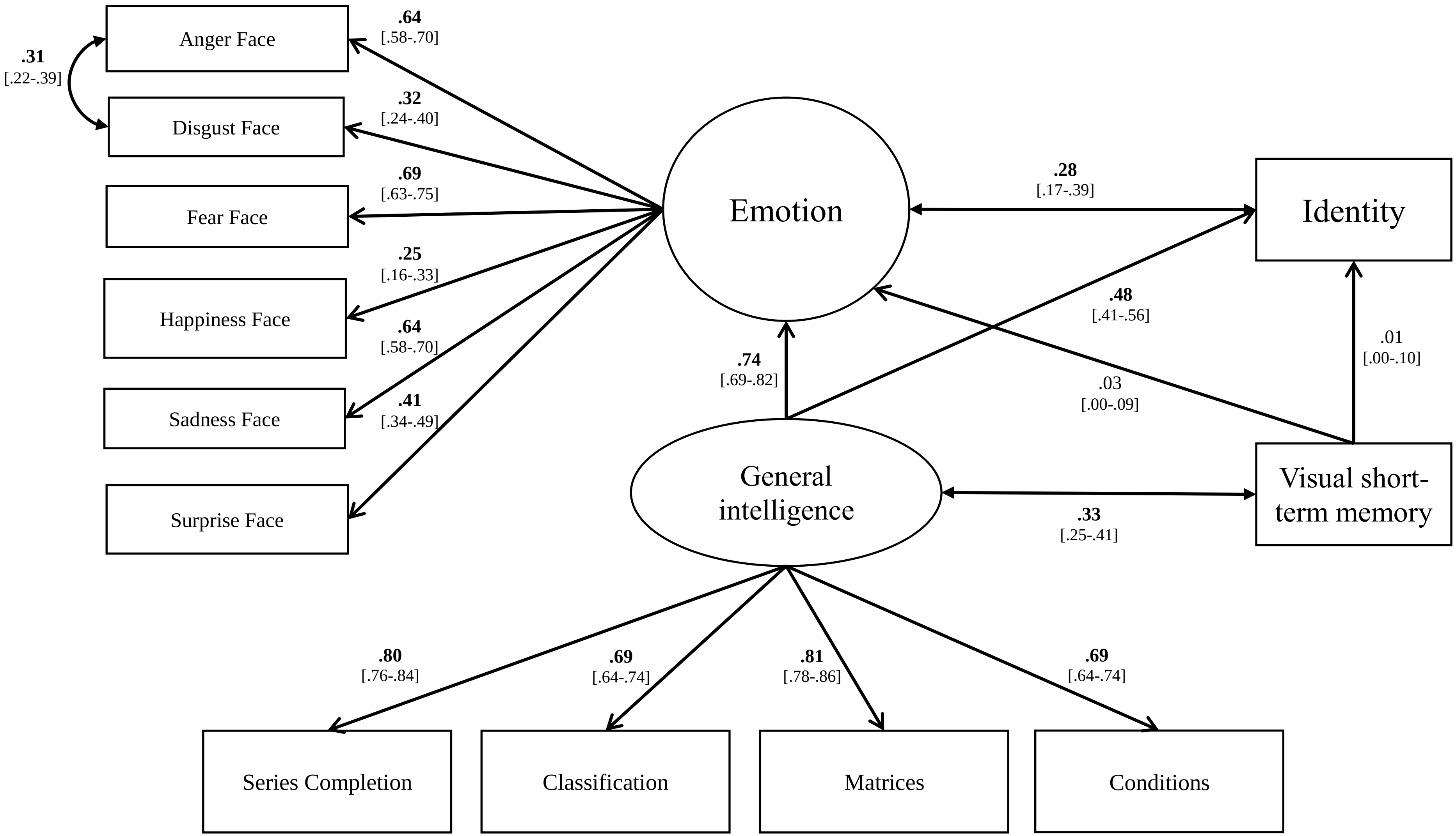
Figure 1. Graphical representation of the first theoretical model with standardized parameter estimates (and 95% confidence intervals) and the added Anger-Disgust pathway. All path coefficients in bold were significant at *p*<.001.



We then fitted a model (including the additional Anger-Disgust parameter as in Figure 1) which included measures of general intelligence and visual short-term memory (VSTM) to assess whether the association between face emotion recognition and face identity recognition was robust to their effects. This model is detailed in Figure 2 and the overall model showed a good fit to the data (CFI: .97, RMSEA: .05). General intelligence was a strong predictor of face emotion recognition (β = .74, *p*<.001) and face identity recognition (β = .48, *p*<.001). In contrast, VSTM was unrelated to these two abilities. Of importance, adjusting for general intelligence and VSTM, the association between face emotion and face identity recognition reduced to *r* = .28, albeit still remaining highly significant (*p*<.001).

As noted in the Methods, we ran our models again exchanging our chosen measure for VSTM with an alternative formulation (i.e. one derived from learning set size 4, rather than the aggregate of learning set sizes 1, 2 and 3), This did not lead to any notable changes in parameter estimates with regard to our core tests; however, the association between general intelligence and VSTM was no longer significant (see Table 2 and Figure 3 in the Supplementary Materials).

Figure 2. Graphical representation of the second theoretical model with standardized parameter estimates (and 95% confidence intervals) and the added Anger-Disgust pathway. All path coefficients in bold were significant at *p*<.001.



4.4 Discussion

In a well-powered and age diverse sample, we observed a large positive association between face emotion recognition and face identity recognition abilities. Importantly, this association – while reduced in magnitude – was still present and highly significant when general intelligence was modelled as a covariate. Moreover, the association remained even when visual short-term memory was modelled as a covariate, consistent with the notion that the overlap reflects face-specific processes.

The results observed here are interesting for a number of reasons. Firstly, the magnitude of the observed association between face emotion recognition and face identity recognition ability is in line - and in fact here largely exceeds - correlations reported in related studies in the field (e.g. Borod et al., 2000; Palermo et al., 2013). Concerns of sample-specific effects and inflated parameter estimates therefore appear to be unwarranted. Secondly, the observation that face emotion recognition and identity recognition ability are associated, over and above the effect of broader cognitive or perceptual processes, is consistent with the idea that that face processing ability reflects different contributions at multiple levels of individual differences that include general cognitive processing, generalised face-perceptual processes, and emotion- and identity-specific processes. This converges both with existing models drawn from other sources of data (Bruce & Young, 1986; Haxby et al., 2000) and with related individual differences research highlighting a hierarchy of abilities (e.g. Lewis et al., 2016). Thirdly, the current findings provide additional support for recent work reporting a general face perception factor '*f*', distinct from other cognitive abilities (Verhallen, Bosten, Goodbourn, Lawrance-Owen, Bargary, & Mollon, 2017; McCaffery, Robertson, Young, & Burton, 2018). This work, however, only assessed identity and other non-expression relevant face tasks. As such, our observations both strengthen the likely existence and broaden the scope of the putative *f* factor. Finally, we observed moderate-to-strong positive correlations between general intelligence and identity recognition ability and emotion recognition ability, respectively. Given the relatively inconsistent reports in the literature of these associations (e.g. Gignac et al., 2016; Lewis et al., 2016; Palermo et al., 2013; Shakeshaft & Plomin, 2015), and the large and diverse sample of the current study, these findings offer a more robust indication of the strength of the association between these two face cognition abilities and general intelligence.

Our use of the Cam-CAN data set restricted us to using the Benton Test of Facial Recognition as our measure of face recognition ability. As a test that involves matching relatively standardised images of unfamiliar faces, this has the limitations that it does not tap key skills of familiar face recognition (Young & Burton, 2017) and the ability to cope with everyday image variability (Bruce, 1994; Burton, 2013). It has also been suggested that the Benton Test allows the use of atypical feature-matching strategies, even by prosopagnosic individuals (Newcombe, 1979; Bowles et al., 2009), though in fact many normal participants rely on feature matching in a variety of matching tasks with unfamiliar faces (Hancock, Bruce, & Burton, 2000). As noted, we observed a strong positive correlation between general intelligence and our measure of face identity recognition ability (the Benton Test of Facial Recognition). This is consistent with the idea of a strategic component and is also of interest because some previous studies did not find a correlation between general intelligence and the related Cambridge Face Memory Test (CFMT: e.g. Wilmer, Germine & Nakayama, 2014) which, as its name implies, has an additional face memory component. One such example is Verhallen and colleagues (2016), who reported no significant correlation between CFMT performance and scores on the standard British school qualification: *General Certificate of Secondary Education* (GCSE).

It is possible that our observed association between face emotion and identity recognition abilities may have been bolstered by the relatively high perceptual demands of both the Benton Test and the face emotion expression recognition test. We note that a weaker relationship may have been observed if a different face identity test had been used that required less contribution from earlier perceptual processes and more contribution from later cognitive stages (as in the case of the CFMT). Nonetheless, we consider it informative that the correlation between our face identity (Benton test) and face emotion (Ekman expressions) recognition tasks remained evident when intelligence was modelled as a covariate.

Some further caveats should be noted. Firstly, our model fit was improved by adding an Anger-Disgust parameter. We had not anticipated a need to do this, but it is in line with modern research that points to a close relation between some forms of anger and disgust (Calder et al., 2010). Indeed, the modern meaning of the word ‘disgust’ seems to be moving away from mere physical revulsion emphasised by Darwin (1872) toward a kind of moral outrage (people say they are disgusted by the greed of bankers, etc.) (Rozin, Haidt, & Fincher, 2009). Secondly, our participants were Western and so we cannot infer that the overlap in face emotion and identity recognition ability necessarily extends to other ethnicities and cultures. Thirdly, we acknowledge that our data here do not make it possible to examine the underlying processes of face emotion and face identity recognition or to detect asymmetric associations between them. We are able only to model covariances between the two abilities, and whilst these are undoubtedly informative, we note that this may be considered a relative limitation of the current design. Finally, it is somewhat difficult to directly compare the association between facial expression and identity recognition that we have observed here with other studies that have measured these abilities using different tasks (e.g. emotion-matching and the CFMT in Palermo et al., 2013). Future studies that can corroborate the pattern we report using a variety of tests and measures will support the stability of these associations.

Conclusion

In summary, we took an individual differences approach to understanding the extent and nature of overlap between emotion and identity recognition abilities. We observed a strong positive correlation between face emotion recognition ability and face identity recognition ability. While this association was reduced in magnitude when adjusting for measures of g and visual short-term memory, it remained moderate and highly significant. These results show a common component involved in face emotion and face identity recognition that is distinct from other cognitive and perceptual mechanisms. We suggest that face processing ability reflects contributions from multiple levels of individual differences, ranging from general cognition to general face processing to emotion-specific and identity-specific processes, and that modelling individual differences can offer an important new perspective on these.

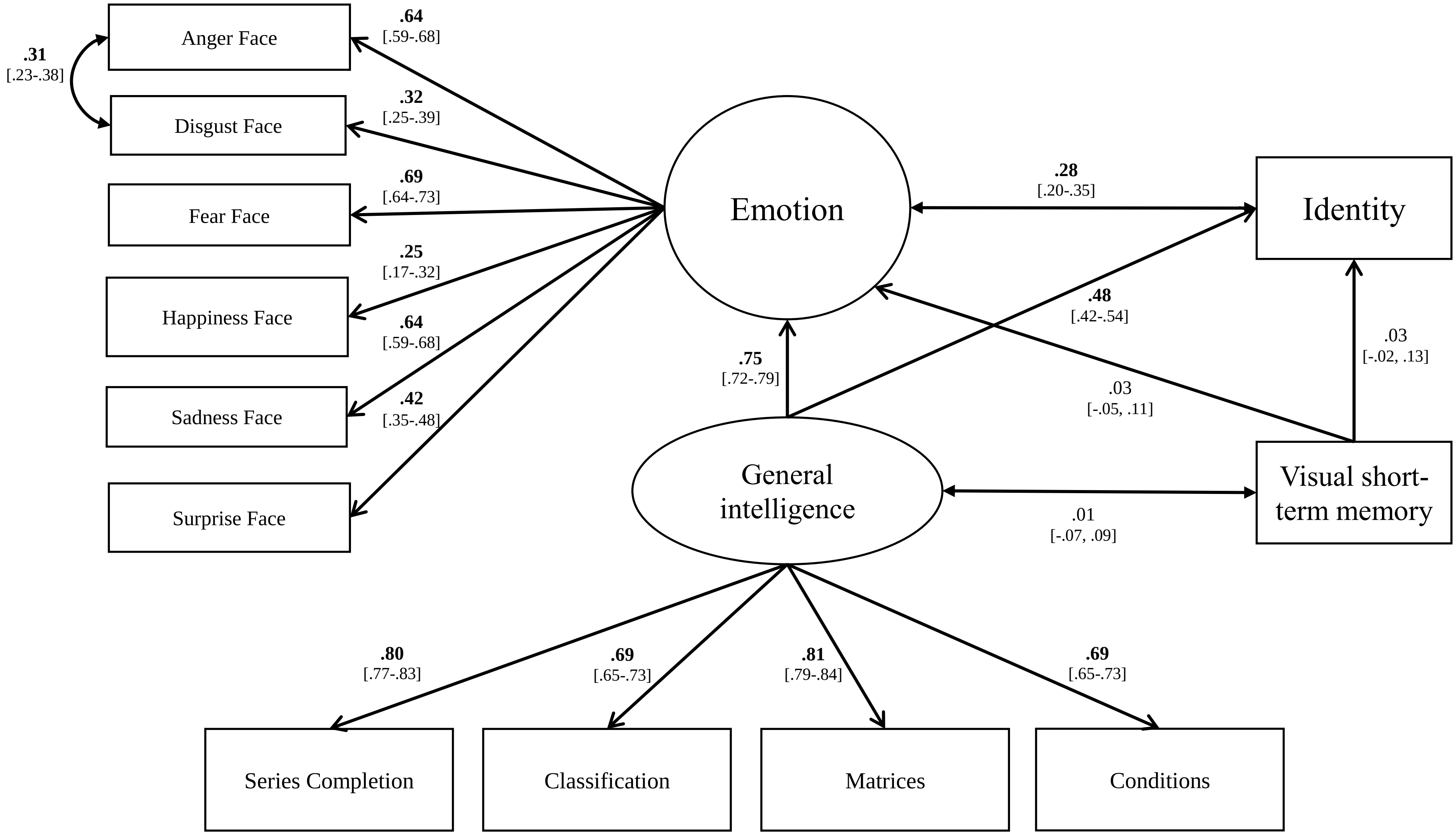
Supplementary Materials

Table 2. Correlations between the emotion recognition scores, the Benton Test of Facial Recognition, the Cattell Fluid Intelligence test and its four subtests, and the Visual Short-term Memory (VSTM) Learning Set Size 4 accuracy score.

	Mean	SD	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Benton	Cattell Total	Series Completion	Classification	Matrices	Conditions
Anger	83.4	19.6												
Disgust	86.2	18.3	**.43**											
Fear	75.3	21.9	**.46**	**.28**										
Happiness	97.5	5.4	**.20**	**.18**	**.19**									
Sadness	92.5	13.3	**.40**	**.23**	**.43**	**.18**								
Surprise	89.5	11.1	**.26**	.12**	**.18**	**.12**	**.31**							
Benton	85.0	8.5	**.30**	.09*	**.38**	.03	**.35**	**.27**						
Cattell Total	69.5	14.6	**.43**	**.15**	**.48**	.12**	**.42**	**.33**	**.42**					
-Series Completion	78.6	16.9	**.39**	**.17**	**.41**	.12**	**.41**	**.34**	**.41**	**.84**				
-Classification	58.1	15.2	**.30**	.06	**.34**	.07	**.29**	**.22**	**.26**	**.81**	**.54**			
-Matrices	75.7	18.2	**.40**	**.15**	**.46**	.13**	**.36**	**.27**	**.43**	**.85**	**.64**	**.58**		
-Conditions	66.2	23.0	**.32**	.10*	**.37**	.08*	**.31**	**.27**	**.29**	**.79**	**.56**	**.53**	**.54**	
VSTM	1.43	.37	.01	-.02	.01	.001	-.07	.03	-.05	.01	-.06	.05	.06	-.01

Note. Performance scores ranged from 5 to 100% accuracy. Note. * $p < .05$; ** $p < .01$; Bold indicates $p < .001$.

Figure 3. Graphical representation of the model with standardized parameter estimates (and 95% confidence intervals) and the added Anger-Disgust pathway and Visual short-term memory (VSTM) derived from Learning Set Size 4. All path coefficients in bold were significant at *p*<.001.



Chapter 5: Alexithymia

Consistent Evidence of a Link between Alexithymia and General Intelligence

The work presented in this section was conducted in collaboration with Andrew Young and Gary Lewis (supervisor) and a revision has been submitted to *Cognition and Emotion.*

Abstract

Alexithymia is a personality construct characterised most notably by a difficulty in identifying and expressing feelings. Although the emotional difficulties in alexithymia are well established, to date little work has examined its relationship to broader cognitive abilities, such as general intelligence. Across three independent healthy adult samples (Ns = 389, 318, & 273), we examined the association of alexithymia to a measure of general intelligence. In all three samples, we observed a significant negative association between alexithymia and general intelligence. In two of the samples, general intelligence was a significant predictor of alexithymia even when accounting for performance on tests of facial expression recognition and of supramodal emotion recognition (measured with faces, bodies, and voices). From a theoretical perspective, these results suggest that models of alexithymia need to incorporate a role for more generalised cognitive functioning. From a practical perspective, studies examining links between alexithymia and clinical disorders, many of which have known links to general intelligence, should consider including a measure of general intelligence in order to adjust for this potentially confounding factor.

5.1 Introduction

Alexithymia is a stable, sub-clinical personality construct characterised by difficulties with emotional awareness, and comprising three components including difficulties in identifying feelings, difficulties in expressing feelings to others, and an externally-oriented style of thinking (Preece, Becerra, Allan, Robinson, & Dandy, 2017). Alexithymia is an important individual differences construct as it has known links to broader psychological difficulties (Mattila, Salminen, Nummi, & Joukamaa, 2006). Indeed, many studies have reported wide-ranging associations with clinically-relevant disorders, such as depression (Grynberg et al., 2012), anxiety (Korkoliakou et al., 2014), eating disorders (Brewer, Cook, Cardi, Treasure, & Bird, 2015), autism spectrum disorder (Cook, Brewer, Shah, & Bird, 2013), atypical empathy (Bernhardt & Singer, 2012), alcohol dependence (Lyvers, Onuoha, Thorberg, & Samios, 2012), and medically unexplained somatic symptoms (De Gucht & Heiser, 2003).

To date, most studies of alexithymia have considered it a deficit in emotion processing. For example, Bird and colleagues (2010) described alexithymia as a “subclinical phenomenon marked by difficulties in identifying and describing feelings and difficulties in distinguishing feelings from the bodily sensations of emotional arousal” (p. 1516). While this description does indeed appear to capture key features of alexithymia, here we sought to address a largely unexamined issue: the possibility that the alexithymia construct might to some extent reflect individual differences in general intelligence.

There are several reasons to posit this broader psychological footprint. Firstly, it has been noted that alexithymia is related to non-emotional interoceptive abilities, with individuals higher in alexithymia showing a poorer ability to accurately count their own heartbeat (Herbert, Herbert, & Pollatos, 2011). A more recent study has bolstered this earlier finding, reporting that alexithymia is negatively correlated with other interoceptive abilities; specifically, the ability to accurately gauge if a drinking solution is saltier than a target solution, to estimate and balance the weights of two-rice filled buckets in each hand, and to gauge the strength of one's own respiratory output (Murphy, Catmur, & Bird, 2018). Such findings indicate that alexithymia cannot be adequately characterised by emotion-related processes alone.

Secondly, several of the items in the most widely-used measure of alexithymia, the Toronto Alexithymia Scale (TAS-20), appear to capture, at least in part, the ability to understand and integrate complex information, which of course is one of the cardinal features of general intelligence (Gottfredson, 1998). For example, agreement with the following items ‘I prefer to analyse problems rather than just describe them’ (reverse-scored) and ‘I prefer to just let things happen rather than to understand why they turned out that way’ both reflect higher levels of alexithymia. Stronger agreement to these items may indeed reflect difficulty with emotional awareness. Equally, it seems plausible that such agreement, at least in part, might reflect difficulties in processing complex contingencies and thus general intelligence.

Thirdly, there are a number of abilities and clinically-relevant disorders that are known to be related both to alexithymia and general intelligence. These include emotion recognition ability in the normal population (Lewis, Lefevre, & Young, 2016) and clinical conditions such as depression (Honkalampi et al., 2000), eating disorders (Lopez, Stahl, & Tchanturia, 2010; Brewer et al., 2015), schizophrenia (Gawęda & Krężołek, 2019), and dementia (Sturm & Levenson, 2011). These again suggest that a link between alexithymia and broader intelligence is plausible.

Only a handful of studies have examined this putative link, and these have mainly relied on measuring focal aspects of intelligence. For example, negative associations have been noted with vocabulary (N=81: Bagby, Taylor, & Ryan, 1986; N=115: Hsing, Mohr, Stansfield, & Preston, 2013). Similarly, Lumley and colleagues (2005) noted a moderate inverse link between verbal ability and the Externally Oriented Thinking subscale of the TAS-20 (N=140:

Lumley, Gustavson, Partridge, & Labouvie-Vief, 2005). In a recent study comprising three samples (Ns=296, 131, and 121), alexithymia was associated with poorer memory for neutral words and poorer executive functioning, as measured by the Trail-making Tests, the Symbol Digit Modalities Test, and a test of category fluency (Correro, Paitel, Byers, & Nielson, 2019). Finally, in a number of large-scale community studies, alexithymia has shown negative associations with the proxy measure of level of education (N=2297: Kauhanen, Kaplan, Julkunen, Wilson, & Salonen, 1993; N=5028: Kokkonen, Karvonen, Veijola, Läksy, & Jokelainen, 2001), consistent with a link to general intelligence (Deary, Strand, Smith, & Fernandes, 2007). However, not all studies have reported a link between alexithymia and intelligence measures. Parker, Taylor, and Bagby (1989) found no association of the TAS-26 scale to the Shipley Institute of Living Scale, which comprises both vocabulary and abstract reasoning subscales.

These studies have thus provided grounds for further examination of how alexithymia might relate to general intelligence; but a number of weaknesses in their study designs necessitate further work. For example, a number of the empirical samples have been small (N≤140) and often limited to university students (e.g. Lumley et al., 2005; Hsing et al., 2013). Many of them use a proxy measure for general intelligence (e.g. Kauhanen et al., 1993, Kokkonen et al., 2001) or use a measure that taps only a limited aspect of intelligence, such as reading ability (Lumley et al., 2005). A comprehensive measure tapping the broad range of intelligence is needed to fully explore a possible link with alexithymia, and may be why Lumley et al. (2005) observed a link between verbal ability and only one of the TAS-20 subscales. Moreover, most of the above studies do not include a good measure of emotion recognition ability (that is, the ability to accurately recognise emotional expressions in other people). Given that emotion recognition has been linked to both intelligence and alexithymia (Lewis et al., 2016), it is important to adjust for this variable to validly assess a link between alexithymia and intelligence. This will enable us to determine whether any association between them is unique or if it is merely reflecting the known association between intelligence and emotion recognition.

With this in mind, we sought here to provide a strong test of the proposed link between alexithymia and general intelligence using well-established measures. Study 1 involved a retrospective analysis of data collected for other purposes from two large adult samples (each with N>300). Study 2 was pre-registered to give a sample (N>270) derived from power calculations based on the effect sizes in Study 1.

5.2 Study 1

In two relatively large adult samples (N>300), we tested whether general intelligence and alexithymia are negatively associated. We also examined the extent to which any such association provides incremental prediction over and above facial expression recognition and supramodal measures of emotion recognition ability (i.e. measures of emotion recognition ability that incorporate stimuli from more than one sensory modality: Study 1a: face and body; Study 1b: face and voice).

5.2.1 Methods

Participants

Study 1a: 400 participants were recruited from Amazon's Mechanical Turk (MTurk) service as part of an unrelated study that included a measure of alexithymia and general intelligence (Lewis et al., 2016). We excluded participants who scored at or below chance level on two or more of the emotion recognition tasks. This resulted in the omission of 11 participants, producing a final sample size of 389 (131 men). Mean age of participants was 37.0 years (SD = 11.7), and ethnicity was reported as follows: White (n=290), Hispanic (n=14), Asian (n=22), Black (n=11), Native American (n=2), Other (n=32) and Undisclosed (n=18).

Study 1b: 412 participants were recruited from Amazon's Mechanical Turk service as part of an unrelated study that included a measure of alexithymia and of general intelligence (Connolly, Lefevre, Young, & Lewis, in press). The exclusion criteria were the same as above, with the addition of participants also being excluded if they failed an auditory test at the beginning of the survey. This resulted in the omission of 94 participants, producing a final sample size of 318 (159 men). Mean age of participants was 35.9 years, and ethnicity was reported as follows: White (n=254), Hispanic (n=21), East Asian (n=14), Black (n=22), Native American (n=2), South Asian (n=1), and Other (n=4). These participant demographics represent the typical pattern observed for MTurk samples (Huff & Tingley, 2015).

Stimuli

Facial Expression Recognition Ability (Face Exp):

Study 1a: Stimuli were static images taken from the Facial Expressions of Emotion: Stimuli and Tests (FEEST) set (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002), and were morphed to differing intensities of expression, from 25% (a morph that involved a 75% neutral and 25% emotional expression) to 100% (representing the prototype expressions from the Ekman series). They were piloted across different levels of emotional intensity in a previous study (Lewis et al., 2016), and in order to try and avoid floor or ceiling effects, exemplars for each emotion with appropriate means and variabilities were selected. In Study 1a, the modality block comprised 50 selected items (10 items for each of the 5 basic emotions: anger, disgust, fear, happiness, and sadness) shown twice to the participants.

Study 1b. For the participants in Study 1b the block comprised an abbreviated subset of stimuli, which were fully outlined in a previous study (Connolly et al., 2020). The subset block comprised 15 items (3 items for each emotion) shown once each.

Alice Heim 4 Test of General Intelligence (AH4):

Study 1a: This test consists of 65 items which are either numerical or verbal in nature. Half of the questions require multiple-choice responses (from a choice of 5), and half require open entry 'creative answers'. The participants have a maximum time limit of 10 minutes to complete as many of the items as they can, after which the page automatically advances to the next section of the survey. They are encouraged to complete the questions in the order in which they are given, but are also told they may skip to the next questions if they become stuck. The participants were awarded one point for every correct answer, and their total score was generated out of a maximum of 65. This test is suitable for the current research question given that it is considered a good test of *general* intelligence as it corporates elements of both crystallized knowledge (vocabulary) and fluid intelligence (online manipulation of novel information). It is also easy and relatively quick to administer online, and relates well to other measures of g (e.g. Tombaugh, 2006).

Study 1b: The identical procedure as for Study 1a was followed.

Toronto Alexithymia Scale (TAS-20): The TAS-20 is a 20-item instrument that assesses alexithymic traits in three broad sub-domains: Difficulty Identifying Feelings (7 items), Difficulty Describing Feelings (5 items), and Externally Oriented Thinking (8 items) (Bagby, Parker, & Taylor, 1994). Participants respond on a 5-point Likert scale, with 1 being 'strongly disagree', and 5 being 'strongly agree'. A total score was generated by summing the responses from all 20 items: a higher score indicates a greater degree of alexithymic traits. Scale scores for each individual sub-scale were also generated. Cronbach's alpha for both datasets was good for the total scale (α=.89; α=.89, respectively), and the Difficulty Identifying Feelings (α=.91; α=.94) and Difficulty Describing Feelings (α=.84; α=.83) sub-scales. The Externally-Oriented Thinking subscale had a relatively low alpha in both samples (α=.67; α=.61), and as such, results of this sub-scale should be interpreted with a degree of caution. Participants in both Studies 1a and 1b completed this test.

Additionally in Study 1a, a measure of body expression recognition ability was included, and in Study 1b, a measure of voice expression recognition ability was included. These are outlined below.

***Point-light Bodily Expression Recognition Ability (Body Exp)*:** Stimuli were dynamic point-light walker stimuli previously described by Atkinson, Dittrich, Gemmell, and Young (2004). In short, 10 actors were recorded performing each of five emotions, and subsequent rendering removed all information other than high visibility patches from each video, resulting in a short video clip of light points whose combined movement simulated a human body expressing a natural, dynamic emotion. Video clips lasted between 4.2 and 8 seconds. As with the facial expression stimuli selected for Study 1a, 10 stimuli for each emotion were chosen (i.e., total N=50) that showed adequate means and variances following a pilot experiment in a previous study (Lewis, Lefevre, & Young, 2016). Participants in Study 1a were shown the 50-item block of bodily expression stimuli twice.

Vocal Expression Recognition Ability (Voice Exp): Stimuli were short audio clips from the Montreal Affective Voices (MAV) set (Belin, Fillion-Bilodeau, & Gosselin, 2008), which comprises actors portraying different emotions in vocal bursts, with each audio burst lasting between 1.45 and 2.23 seconds. To ensure that this set was in line with the other emotion stimuli, it was first piloted in a previous study (Connolly et al., 2020) to select appropriate items for the five basic emotions, and then to select 15 items (3 for each emotion) with psychometric properties comparable to the other two expressive domains. Participants in Study 1b were presented with the 15-item block of vocal expression stimuli once.

Procedure

In both samples, the emotion expression recognition blocks were first presented to the participant (faces and bodies in Study 1a, faces and voices in Study 1b), with the within-block presentation order being fully randomised. In a five alternative-choice paradigm, participants had to select the emotion they perceived was being portrayed. Stimuli presentation consisted of a black screen for 500ms, a fixation cross for 750ms, and a further 500ms black screen that preceded the onset of the stimulus. Facial stimuli in both samples were presented for 1000ms. Bodily (Study 1a) and vocal (Study 1b) stimuli lasted for the duration of each individual video or audio clip. Participants could respond at any point following the stimulus onset. Following the emotion expression recognition blocks, participants completed as much of the general intelligence test as possible in the maximum time of 10 minutes allowed. Following this, they completed the alexithymia questionnaire. Participants were debriefed following completion.

5.2.2 Results

A similar pattern of associations was noted across both samples. First, we explored links to facial expression recognition, as this has been a focus in previous theorising. Alexithymia was significantly negatively correlated with general intelligence (Study 1a: r=-.11, p=.026; Study 1b: r=-.38, p<.001) and with facial expression recognition ability (Study 1a: r=-.28, p<.001; Study 1b: r=-.35, p<.001). Additionally, facial expression recognition was significantly positively correlated with general intelligence (Study 1a: r=.28, p<.001; Study 1b: r=.38, p<.001).

For completeness, we then investigated links to emotion recognition more generally, by examining the correlations when emotion recognition was based on a composite supramodal score derived from the face and body expression recognition scores (Study 1a) or face and voice expression recognition scores (Study 1b). Alexithymia was significantly negatively correlated with supramodal expression recognition ability (Study 1a: r=-.18, p=.003; Study 1b: r=-.36, p<.001). Additionally, supramodal expression recognition was significantly positively correlated with general intelligence (Study 1a: r=.38, p<.001; Study 1b: r=.45, p<.001).

Hierarchical regression analysis then enabled our other hypotheses to be tested. In the first step, age and sex were entered as predictors of alexithymia. In Study 1a, neither of these variables were significant predictors. In Study 1b, age was a significant negative predictor of alexithymia (β=-.11, p=.046), whereas sex was not. In the second step, facial expression recognition ability was entered into the respective models. In both samples, this was a significant negative predictor of alexithymia (Study1a: β=-.27, p<.001; Study 1b: β=-.34, p<.001). In Study 1a, this was the only significant predictor, but in Study 1b, age also remained a significant negative predictor (β=-.12, p=.031).

In the critical third step, general intelligence was entered into the model. In Study 1a, this variable was not a significant predictor of alexithymia, but facial expression recognition remained a significant negative predictor (β=-.26, p<.001). Including intelligence in the model did not improve the overall fit (R^2 changed from .081 to .082, p=.560). In Study 1b, however, both general intelligence (β=-.29, p<.001) and facial expression recognition ability (β=-.23, p<.001) were significant negative predictors of alexithymia, and the inclusion of general intelligence significantly improved the fit of the model (R^2 changed from .135 to .205, p<.001). The full results of both regression analyses are shown in Tables 1 (Study 1a) and 2 (Study 1b). Note that the coefficients for sex reflect alexithymia scores being higher for women in Study 1a, and higher for males in Study 1b, although neither of these reach statistical significance.

Table 1: Multiple regressions predicting alexithymia scores from age, sex, facial expression recognition ability (Face Exp), and general intelligence in Study 1a.

	Step 1		Step 2		Step 3	
Independent Variables	**β**	**Sig**	**β**	**Sig**	**β**	**Sig**
Age	-.08	.119	-.05	.302	-.05	.285
Sex	.09	.068	.04	.439	.04	.413
Face Exp			**-.27**	**<.001**	**-.26**	**<.001**
General Intelligence					-.03	.560

Table 2: Multiple regressions predicting alexithymia scores from age, sex, facial expression recognition ability (Face Exp), and general intelligence in Study 1b.

	Step 1		Step 2		Step 3	
Independent Variables	**β**	**Sig**	**β**	**Sig**	**β**	**Sig**
Age	**-.11**	**.046**	**-.12**	**.031**	-.10	.059
Sex	-.10	.081	-.04	.513	-.04	.406
Face Exp			**-.34**	**<.001**	**-.23**	**<.001**
General Intelligence					**-.29**	**<.001**

Again, for completeness, we ran the regression models with ‘supramodal’ emotion recognition scores derived from a composite of either face and body expression recognition (Study 1a) or face and voice expression recognition (Study 1b). The results were highly similar to the models that tested face expression recognition alone. That is, they followed the exact pattern as in the outputs in Tables 3 and 4: the magnitude of the beta values were not appreciably different and the same predictor variables at each step were significant as in the facial expression regression models.

5.2.3 Discussion

We observed the expected negative association between alexithymia and emotional expression recognition in both samples. Importantly, we also observed a significant negative correlation between alexithymia and general intelligence across samples. Of note, though, in Study 1a, general intelligence was not a significant predictor of alexithymia over and above that of emotion recognition. In contrast, in Study 1b, both general intelligence and emotion recognition were significant predictors when entered in the model simultaneously. These patterns held regardless of whether emotion recognition was measured from facial expressions alone or from supramodal performance involving faces and voices or bodies.

These findings clearly leave unanswered the question as to whether intelligence has a *unique* association with alexithymia once expression recognition ability is statistically adjusted for. In order to provide a more definitive answer to this research question, we addressed this issue using an independent sample and a pre-registered study plan.

5.3 Study 2

5.3.1 Methods

Participants

This pre-registered study (https://osf.io/67zda/) sought to provide 80% power with an alpha level of .05 to detect an increase in R^2 of .03. It also sought to provide 80% power with an alpha of .05 (one-tailed) to test for a zero-order association of $r \geq .15$. These effect sizes allowed us to provide a strong test of the results observed in Study 1b. A power analysis indicated that an N of 270 was required to achieve sufficient power under these conditions and so we recruited participants until the desired N was reached after removing those who met our exclusion criteria. Therefore, a total of 357 participants were recruited from Amazon's Mechanical Turk service. A proportion of participants failed the pre-survey auditory test, and were excluded before completing the survey (N=25). We also excluded participants who showed chance level performance on at least two of the emotion recognition batteries (N=53). Finally, a small number of participants indicated that they had not completed the survey seriously and were thus excluded (N=6). This resulted in an omission of 84 participants, producing a final sample size of 273 (138 men). Participants gave informed consent and ethical approval was granted by the Royal Holloway Ethics Committee.

Mean age of participants was 38.5 years (SD = 12.2, Range=19-74), and ethnicity was reported as follows: White (n=192), Hispanic (n=18), Asian (n=18), Black (n=36), Native American (n=1), Middle Eastern (n=1), Other (n=6) and Undisclosed (n=1). As with the previous two samples, these participants represent the typical demographic pattern for MTurk samples (Huff & Tingley, 2015).

Stimuli

Facial Expression Recognition Ability (Face Exp): The same facial expression battery was used as in Study 1b; i.e. a subset of items from Study 1a taken from the FEEST set (Young et al., 2002). The subset block comprised 15 items (3 items for each basic emotion) that were shown to participants once each.

Bodily Expression Recognition Ability (Body Exp)**:** Bodily emotional stimuli were taken from the same battery used in Study 1a (Atkinson et al., 2004). In line with the facial and vocal batteries, a subset of these 50 items were piloted in a previous study to ensure suitability for individual differences research, and 15 items with appropriate psychometric properties were selected. This abbreviated battery comprised 15 items (3 items for each basic emotion) shown to participants once.

Vocal Expression Recognition Ability (Voice Exp): The same vocal stimuli battery was used as in Study 1b (Belin et al., 2008). The battery comprised 15 items (3 items for each basic emotion) presented to participants once.

Alice Heim 4 Test of General Intelligence (AH4): This same 65-item general intelligence was used as in Studies 1a and 1b.

Toronto Alexithymia Scale (TAS-20): The same 20-item alexithymia measure was used as in Studies 1a and 1b. Cronbach's alpha for our data was good for the total scale (α=.91) and for the Difficulty Identifying Feelings (α=.94) and Difficulty Describing Feelings (α=.83) sub-scales. As with the previous two samples, the alpha for the Externally-Oriented Thinking subscale was fairly low (α=.67). Low reliability for this subscale has been noted by several other researchers in the field (e.g. Meganck, Vanheule, & Desmet, 2008; Preece et al., 2017) and future revision of this scale has been suggested.

Procedure

The emotion expression recognition blocks were first presented to the participant in the same fixed order (voice, face, body), with the within-block presentation order being fully randomised. Stimulus presentation duration and response format were identical to the previous samples. Following the emotional expression recognition blocks, participants completed as many items on the general intelligence test as they were able to in 10 minutes. Finally, they completed the alexithymia questionnaire. Participants were debriefed following survey completion.

Analysis

The same analyses were conducted as in Studies 1a and 1b, but with one difference. In both Studies 1a and 1b, expression recognition ability comprised either facial expression stimuli alone, or a composite supramodal score across two different expressive modalities (faces and bodies in Study 1a; faces and voices in Study 1b). In Study 2, all participants were tested on facial, bodily, and vocal expression recognition, and therefore, expression recognition ability was measured as both facial (in line with the previous two samples) and as a trimodal supramodal score (derived from facial, vocal, and bodily emotion recognition, and intended as a more ecologically generalisable measure). Therefore, each of the following analyses was carried out twice, with either facial or supramodal emotion recognition measures, in line with our pre-registered protocol.

Firstly, we assessed the association between our variables of interest, in order to answer our first research question of whether alexithymia is negatively correlated with intelligence and (facial or supramodal) expression recognition ability. Secondly, hierarchical regression analyses were used to examine the unique associations of general intelligence and emotional expression recognition ability to alexithymia. As in the previous two samples, we conducted a three-step multiple regression analysis with alexithymia as the dependent variable.

5.3.2 Results

Pre-registered analyses

As seen in Study 1, alexithymia was significantly negatively correlated with both general intelligence (r=-.42, p<.001) and with facial expression recognition ability (r=-.22, p<.001). Facial expression recognition was significantly positively correlated with general intelligence (r=.35, p<.001).

A very similar pattern of results was also observed for supramodal emotion recognition ability. Alexithymia was significantly negatively correlated with supramodal emotion recognition ability (r=-.35, p<.001). Supramodal emotional expression recognition was significantly positively correlated with general intelligence (r=.51, p<.001).

We next used hierarchical linear regression to test whether general intelligence was a predictor of alexithymia over and above emotion recognition ability (and other variables). In the first of these models, age and sex were entered as predictors of alexithymia in the first step. Neither of these variables reached statistical significance. In the second step, facial expression recognition ability was entered, which was a significant negative predictor of alexithymia (β=-.20, p=.001). In the third step, general intelligence was entered into the model. This variable was a significant predictor of alexithymia (β=-.40, p<.001), but facial expression recognition was then no longer significant. Inclusion of general intelligence in the model significantly improved model fit (R^2 changed from .054 to .191, p<.001). The full results of the regression analysis are shown in Table 3.

Table 3: Multiple regressions predicting alexithymia scores from age, sex, facial expression recognition ability (Face Exp), and general intelligence in Study 2.

	Step 1		Step 2		Step 3	
Independent Variables	**β**	**Sig**	**β**	**Sig**	**β**	**Sig**
Age	-.03	.614	-.03	.672	-.03	.576
Sex	-.11	.068	-.08	.209	-.10	.094
Face Exp			**-.20**	**.001**	-.06	.317
General Intelligence					**-.40**	**<.001**

In the second of these models, age and sex were first entered as predictors of alexithymia: the results were of course identical to those above. In the second step, supramodal emotion recognition ability was entered and was observed to be a significant negative predictor of alexithymia (β=-.34, p<.001). In the third step, general intelligence was entered into the model. General intelligence was a significant predictor of alexithymia (β=-.33, p<.001), and supramodal expression recognition also remained a significant negative predictor (β=-.17, p=.008). The full results of this regression analysis are shown in Table 4.

Table 4: Multiple regressions predicting alexithymia scores from age, sex, supramodal emotion recognition ability (Supra Emotion), and general intelligence in Study 2.

	Step 1		Step 2		Step 3	
Independent Variables	**β**	**Sig**	**β**	**Sig**	**β**	**Sig**
Age	-.03	.614	-.05	.391	-.04	.449
Sex	-.11	.068	-.06	.324	-.08	.156
Supra Emotion			**-.34**	**<.001**	**-.17**	**.008**
General Intelligence					**-.33**	**<.001**

Subsidiary Exploratory Analyses

As an exploratory step to complement the pre-registered analyses, we examined the correlations between general intelligence and the three alexithymia subscales across all three of our sets of data.

In Study 1a, only the Externally Oriented Thinking (EOT) subscale showed a significant negative association with intelligence: r=-.13, p=.014). Neither the Difficulty Identifying Feelings (DIF) nor the Difficulty Describing Feelings (DDF) subscales showed significant associations with intelligence (DIF: r=-.09, p=.077; DDF: r=-.06, p=.251). In Study 1b, all three of the subscales were significantly and negatively associated with intelligence (DIF: r=-.38, p<.001; DDF: r=-.25, p<.001; EOT: r=-.26, p<.001).

In Study 2, all three of the alexithymia subscales were significantly and negatively correlated with intelligence (DIF: r=-.38, p<.001; DDF: r=-.31, p<.001; EOT: r=-.34, p<.001).

The results from across the three samples therefore suggest that the association between total alexithymia score and general intelligence is not being solely driven by any particularly cognitively oriented items, for example, those in the Externally Oriented Thinking subscale.

Secondly, in light of an interesting recently proposed perspective, involving the idea of verbal difficulty giving rise to the alexithymia construct, we carried out some exploratory post-hoc analyses that were also not included in our original pre-registration plan. Hobson and colleagues (2019) recently proposed the alexithymia-language hypothesis positing that verbal impairment may be one of several mechanisms underpinning the aetiology of the alexithymic phenotype. This idea arose from the finding that language processes are widely considered to contribute to typical emotional development (Ornaghi & Grazzani, 2013). Indeed, Hobson et al. (2019) report evidence from clinical populations in which early speech and language difficulties are associated with emotion processing difficulties in later life, for example in deaf or autistic children, and cite alexithymia as a key mediator between linguistic and emotional capabilities (Hobson, Brewer, Catmur, & Bird, 2019).

Since our measure of general intelligence included both verbal and nonverbal components, we sought to provide a test of this alexithymia-language hypothesis by assessing whether the verbal intelligence sub-score was more predictive of alexithymia than nonverbal intelligence. In Study 1a, the correlations between the verbal and nonverbal components with alexithymia were -.11 (p=.039) and -.11 (p=.028), respectively, which did not differ significantly from each other (p=.50). In Study 1b, the correlations between the verbal and nonverbal components with alexithymia were -.40 (p<.001) and -.33 (p<.001), respectively, which did significantly differ from each other (p=.023). In Study 2, the correlations between the verbal and nonverbal components with alexithymia were -.43 (p<.001) and -.37 (p<.001), respectively, and whilst these did not statistically differ from each other (p=.053), interpreting p values close to the nominal threshold should be done with caution. Overall, then, these results suggest that alexithymia is associated with general intelligence and that this association is not driven by a specific relationship with verbal intelligence. However, we discuss this interpretation in more depth below.

5.4 General Discussion

The construct of alexithymia as an impairment in emotional understanding has been studied in some depth, especially in clinical populations. To date, however, little work has examined the relation of alexithymia to broader cognitive abilities. In the current study, across three independent samples, we consistently observed a significant negative association between alexithymia and general intelligence, and the magnitude of this association was moderate-to-large in two of our three samples.

Regarding the question of whether alexithymia and intelligence have an independent association, over and above the effects of emotion recognition ability, the results were somewhat more mixed. However, on the balance of evidence – that is, the results from Study 1b and our pre-registered Study 2 – we conclude that general intelligence is uniquely associated with alexithymia, over and above emotion recognition ability. This finding is important because it provides consistent evidence that the psychological footprint of alexithymia extends beyond affect/emotion-relevant processes and abilities, and that a full understanding of individual differences in alexithymia will likely require including a role for general intelligence.

In addition to our core research questions, data from the current study enabled us also to test a recent hypothesis in the literature that posited a specific link between verbal intelligence and alexithymia (Hobson et al., 2019). Across the three samples, we observed relatively limited evidence in favour of this hypothesis. In Study 1a, we observed a nominally significant difference between the verbal and non-verbal associations with alexithymia in the predicted direction. In Study 1b, we did not see this difference. In Study 2, we observed a 'marginally' significant difference in the predicted direction. These mixed results, along with the modest observed effect sizes, indicate that verbal intelligence does not seem to be a special relative of alexithymia, in contrast to the model proposed by Hobson et al. (2019).

What might underlie the observed association between alexithymia and general intelligence? There are several proposed aetiologies of alexithymia, one of them concerning difficulties in evaluating an emotional response, as posited by the attention-appraisal model (Preece et al., 2017). In this model, it is proposed that alexithymic individuals have under-developed emotion schemas, and as such, they experience an emotional response but do not attend to it or evaluate it appropriately, leading to difficulties such as poor emotional regulation. Given that the development of emotion schemas depends on sufficient cognitive development

(Izard, 2011), it is plausible that poor development of cognitive structures may underlie the difficulties in both emotion recognition and general intelligence. For example, if an individual is poor at evaluating complex information, this would have an adverse effect on their general cognitive ability and their recognition of emotional stimuli.

Our findings have some practical implications for examining alexithymia in clinical disorders that are known to show an association with general intelligence, including schizophrenia (Fett, Viechtbauer, Penn, van Os, & Krabbendam, 2011) and depression (Marazziti, Consoli, Picchetti, Carlini, & Faravelli, 2010). In these populations, what may appear to be alexithymic traits may actually reflect deficits in broader cognitive ability, and given the associations reported here, it will be important in future to adjust for general intelligence to eliminate this potential confound.

Some potential limitations of the current work require mention. Firstly, the studies reported here were carried out online on a commonly used data collection platform: Amazon's Mechanical Turk (MTurk). Although there have been concerns raised about the quality of online data, other studies have suggested that they provide reliable, replicable results in a cost- and time-efficient way (Miller, Crowe, Weiss, Maples-Keller, & Lynam, 2017). Indeed, there may be benefits to online testing over traditional lab-based student samples; for example larger, more diverse samples that are more representative of the general US population and who have high internal motivation to respond honestly and accurately (Gosling, Vazire, Srivastava, & John, 2004). Our choice of the MTurk platform is bolstered by the extensive use of it in previous work (e.g. Ramsey, Thompson, McKenzie, & Rosenbaum, 2016; Mortensen & Hughes, 2018).

Secondly, while we make a case here that alexithymia is associated with general intelligence, over and above emotion recognition ability, it is conceivable that this association would be better characterised at a more focal level: that is, our measure of general intelligence may simply be a proxy measure of a construct like interoceptive ability or some kind of general social perception ability. Although this perspective does not invalidate our core argument – namely, that alexithymia clearly extends beyond the emotion sphere itself - further work will now be required to examine this possibility and thus further refine our understanding of the alexithymia construct.

Finally, future work should build on the association we report by including a measure of mood. Recent evidence suggests that the DIF subscale of the TAS-20 captures an individual's distress level rather than their alexithymia (Preece et al., 2020). Given also that there exists an association between cognitive ability and depression (Rock, Roiser, Riedel, & Blackwell, 2014), it would be appropriate to adjust for a measure of negative affect to assess if and how the association between alexithymia and intelligence changes with the inclusion of this potential confound.

Conclusion

In summary, across three independent healthy samples, we observed significant negative associations between alexithymia and general intelligence. Importantly, in two of the samples, intelligence was a significant predictor of alexithymia even after adjusting for emotional expression recognition ability. These results suggest that to achieve a full understanding of the alexithymia construct, theorists will need to engage with broader cognitive abilities and processes.

Chapter 6: Sex Differences in Emotion Recognition

Sex differences in emotion recognition: evidence for a small overall female superiority on facial disgust

The work presented in this section was conducted in collaboration with Carmen Lefevre, Andrew Young, and Gary Lewis (supervisor) and is published in *Emotion.*

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2019). Sex differences in emotion recognition: Evidence for a small overall female superiority on facial disgust. *Emotion*, *19*(3), 455-464.

https://doi.org/10.1037/emo0000446

Abstract

Although it is widely believed that females outperform males in the ability to recognise other people's emotions, this conclusion is not well supported by the extant literature. The current study sought to provide a strong test of the female superiority hypothesis by investigating sex differences in emotion recognition for five basic emotions using stimuli well-calibrated for individual differences assessment, across two expressive domains (face and body), and in a large sample (N=1022: Study 1). We also assessed the stability and generalisability of our findings with two independent replication samples (N=303: Study 2, N=634: Study 3). In Study 1, we observed that females were superior to males in recognising facial disgust and sadness. In contrast, males were superior to females in recognising bodily happiness. The female superiority for recognition of facial disgust was replicated in Studies 2 and 3, and this observation also extended to an independent stimulus set in Study 2. No other sex differences were stable across studies. These findings provide evidence for the presence of sex differences in emotion recognition ability, but show that these differences are modest in magnitude and appear to be limited to facial disgust. We discuss whether this sex difference may reflect human evolutionary imperatives concerning reproductive fitness and child-care.

6.1 Introduction

The ability to accurately recognise other people's emotions is a core socio-cognitive skill (Bruce & Young, 1986, 2012; Young, 2016). Although much work in this domain has emphasised that emotion recognition ability is to some extent akin to an innate human faculty (Darwin, 1872), such that certain emotions are universally recognised, regardless of culture (Ekman & Friesen, 1971), it is also clear that not all people can recognise emotional expressions equally well (Lewis, Lefevre, & Young, 2016). With regard to these individual differences, one of the most widely discussed factors that may influence emotion recognition ability is biological sex (Kret & de Gelder, 2012), where meta-analytic work has claimed that on average women outperform men (Hall, 1978; McClure, 2000; Thompson & Voyer, 2014).

However, despite the widely-held assertion of female ‘superiority’ in emotion recognition ability, a closer examination of the literature reveals a more mixed picture, as detailed below. Furthermore, the literature is almost entirely reliant on studies of facial expression recognition. In other expressive domains, such as recognition of emotion from body postures, little is known regarding sex differences. This warrants further work, as emotion recognition ability has important real-world implications: accurate recognition of emotions is associated with better social functioning (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006), whereas recognition difficulties are linked to interpersonal problems and to the aetiology and maintenance of several psychiatric disorders, such as depression (Surguladze et al., 2004). Given the different prevalence rates of psychopathology across the sexes, sex differences in basic socio-cognitive processes (such as emotion recognition ability) which may contribute to such outcomes (e.g. in depression: Harmer, Goodwin, & Cowen, 2009) are thus important phenomena to delineate and study. Moreover, different evolutionary and socialisation niches occupied by males and females make the study of sex differences important for basic scientific enquiry.

Sex Differences in Facial Emotion Recognition – A Brief Overview

Much of the earlier work in the field has investigated emotion recognition as a general ability based on a global score rather than distinguishing between individual emotions (e.g. Buck, Miller, & Caul, 1974; Kirouac & Doré, 1985). More recent work instead tests recognition of affect more specifically through the use of a wider range of individually-scored basic emotions. In line with this approach, here we explicitly focus on recent research examining the five basic emotions of anger, disgust, fear, happiness, and sadness.

In a study addressing recognition ability for negatively-valenced basic emotions (anger, disgust, fear, and sadness), Rotter and Rotter (1988) (N_{Study1}=727: 214 males; N_{Study2}=399: 162 males) noted a female advantage across all expressions. Hall and Matsumoto (2004), across two studies with different stimuli exposure times (Study 1: 10s, Study 2: 200ms) and using five basic emotions, found that females were significantly more accurate for disgust, happiness, and sadness (Study 1: N=96: 69 males), and for anger, disgust, fear, and sadness (Study 2: N=363: 126 males). Lee and colleagues (2013), using a large sample of adolescents (N=1954: 956 males), found a female advantage for discriminating morphed facial expressions on the continua of Happiness-Fear and Happiness-Sadness (but not the Anger-Sadness or Anger-Fear continua). More recently, Duesenberg and colleagues (2016) tested 80 participants (40 males) on two emotions, anger and sadness, each presented at two intensities of expression, 40% and 80%. These authors reported no sex differences for the sadness stimuli, but found a significant female advantage for accurately identifying anger at both intensities.

However, in contrast to these studies that were broadly supportive of a female advantage in emotion recognition, a number of studies have reported either no sex differences or even a male advantage. Rahman, Wilson and Abrahams (2004) assessed 240 participants (120 males) on happiness and sadness, but reported no accuracy advantage for either emotion. Grimshaw, Bulman-Fleming, and Ngo (2004) (N=73: 36 males) presented stimuli for 50ms portraying three basic emotions of anger, happiness, and sadness, but found no sex difference in accuracy or reaction times. Unlike many other studies in this area which used a forced choice paradigm, Williams and Mattingley (2006) tested 156 (78 males) participants on rapid detection of angry or fearful faces amongst neutral distractors. Males were found to be significantly faster at identifying angry male faces, but there was no difference in speed between sexes to detect fearful stimuli. Again using a visual search paradigm, Sawada and colleagues (2014) (N=90: 46 males) measured participant ability in detecting either an angry or happy face amongst other neutral faces, and found no significant effect of sex on reaction time. Testing a large sample of undergraduates (N=993: 211 males) on facial expressions in both frontal and profile views, Matsumoto and Hwang (2011) did not observe significant sex differences in ability to categorise the five basic emotions. Most recently, Lyusin and Ovsyannikova (2016) tested a large sample of participants (N=684: 221 males) on recognition accuracy and sensitivity on 15 different emotions (including the five basic emotions) using naturalistic video recordings, but reported no significant difference between the sexes on these measures.

Sex Differences in Bodily Emotion Recognition – A Brief Overview

Bodily emotion recognition is the ability to distinguish a portrayed emotion from the stimulus' body form alone. It is a fundamental component of accurate emotion perception, with research showing that body posture can be critical in resolving ambiguous facial expressions (Aviezer, Trope, & Todorov, 2012). The ability to accurately perceive emotion from both face and body may therefore represent an optimal strategy for emotion recognition, in that all the available cues are integrated to form the most informed interpretation (Young & Bruce, 2011; Young, 2018). This strategy seems especially pertinent in situations in which one channel is occluded or ambiguous, or when moving face and body signals are expressed very rapidly.

Body emotion recognition ability can be assessed through both static and dynamic stimuli. In the case of static stimuli, a photograph of a person with their face occluded and expressing a given emotion is presented to the participant. Dynamic stimuli often consist of point-light displays showing a set of locations on a human body making a natural movement (see Johansson, 1973, and Figure 1, or the video in the Supplementary Materials). Whereas a large body of research has addressed sex differences in facial emotion recognition, only a handful of studies to date have addressed sex differences in body emotion recognition. An early study by Sogon and Izard (1987) (N=94: 47 males) involved short video clips of scenes portraying five emotions (surprise, contempt, affection, anticipation, or acceptance), and found females to be significantly better at identifying disgust and fear. Additionally their ability to recognise sadness also approached significance (Sogon & Izard, 1987). A more recent study (N=37: 15 males) involving happy, sad, and angry body stimuli found females to be faster in recognising these emotions from point-light displays, but found no significant sex difference in overall accuracy rates (Alaerts, Nackaerts, Meyns, Swinnen, & Wenderoth, 2011). Two further studies used happy and angry point-light displays, and observed a significant superiority for males in identifying happiness, as well as a tendency, albeit non-significant, for females to perform better at identifying anger (Sokolov et al., 2011; Krüger et al., 2013). However, it should be noted that all of these more recent studies used small samples (N<100) and assessed only two (Sokolov, 2011; Krüger, 2013) or three (Alaerts et al., 2011) of the basic emotions.

The Current Study

To summarise the literature, a wide variety of claims have been made regarding sex differences in face and body emotion recognition ability. Despite meta-analytic work reporting an overall female advantage (Hall, 1978; McClure, 2000; Thompson & Voyer, 2014), a number of studies – some with relatively large samples (e.g. Lyusin & Ovsyannikova, 2016; Matsumoto & Hwang, 2011) – have not reported this pattern. Various studies reported sex differences only for specific emotion(s), and limited consensus can be reached given the nature of the extant literature.

Several factors may explain these mixed findings. Firstly, sample sizes have tended to be modest (N<200) and as such may be underpowered to detect what are unlikely to be large effect sizes. For example, Thompson and Voyer (2014) reported a Cohen's *d* of .19 in favour of females on emotion recognition tasks and a suitably powered test for an effect of this magnitude would require several hundred participants. Secondly, methodological factors may play a role, as widely used stimulus sets have often been developed with a view to creating easily recognised expressions. In consequence, a number of published studies show clear ceiling effects for their stimuli, which will diminish the possibility of detecting group differences (e.g. Hampson, van Anders, & Mullin, 2006; Hoffmann, Kessler, Eppel, Rukavina, & Traue, 2010). Thirdly, while a recent meta-analysis reported evidence for a generalised sex difference in emotion recognition ability (Thompson & Voyer, 2014), evidence of publication bias (as assessed by the Test of Excess Significance method proposed by Ioannidis & Trikalinos, 2007) was apparent. Finally, in the specific case of body emotion recognition, too few studies have been performed to gain traction on possible sex differences.

The present work therefore examined whether sex difference are present in emotion recognition ability, and if so, whether this sex difference is restricted to a specific emotion (across five basic emotions) or expressive domain (i.e., face or body stimuli). With the aforementioned issues in mind, our study made a contribution to the literature in four important ways. Firstly, we used a large sample of adults (N=1022: Study 1) in order to provide adequate statistical power to detect even modest group differences. Second, we used carefully developed stimulus sets – piloted prior to the current investigation – that do not show floor or ceiling effects and thus enhance the power to detect group differences (see Methods and Lewis et al., 2016 for further details). Thirdly, we used both face and body stimuli in order to examine if sex differences are restricted to one expressive domain or instead reflect more general processes. Finally, and with recent discussions of reproducibility of some findings in psychology (Open Science Collaboration, 2015) very much in mind, we used two further independent participant datasets (N=303: Study 2; N=634: Study 3) and additional different sets of face and body stimuli (Study 2) to test the robustness of any effects observed.

6.2 Study 1

6.2.1 Methods

Participants

1063 participants were recruited through Amazon's MTurk service as part of a previous study that did not analyse sex differences in the obtained data (Lewis et al., 2016: Studies 1 and 2). Ethical approval for this study was granted by the Department of Psychology Ethics Committee at the University of York. As expected with an online presentation, a number of participants experienced technical failures (e.g. stimuli not displaying properly). Consequently, and in line with our previous data exclusion strategy we only included participants in our analyses who completed at least 90% (≥18 of 20) of trial blocks for each emotion (anger, disgust, fear, happiness, sadness) and expressive domain (face and body). We also excluded participants for whom responses indicated low attention (e.g. using the same response key repeatedly) and those who did not disclose their sex (N=7). This led to the omission of 41 participants and a final sample size of 1022 (322 males). The gender ratio was near-identical across White and non-White participants. In order to detect the effect size reported in the most recent meta-analysis (Thompson & Voyer, 2014), our sample size provided power of .80 for a Cohen's *d* of .19 with a two-tailed t-test and an alpha level of .05 (Faul, Erdfelder, Lang, & Buchner, 2007). Note, this power analysis reflects an allocation ratio of 2.2: i.e. there were just over twice as many females than males in this sample. The mean age of the sample was 36.2 years (SD = 12.0). A range of ethnicities were reported in the final sample: White (n=775), Hispanic (n=47), Asian (n=53), Black (n=30), Native American (n=11), Other (n=74), and Undisclosed (n=32). These demographics are typical for MTurk samples (Paolacci, Chandler, & Ipeirotis, 2010).

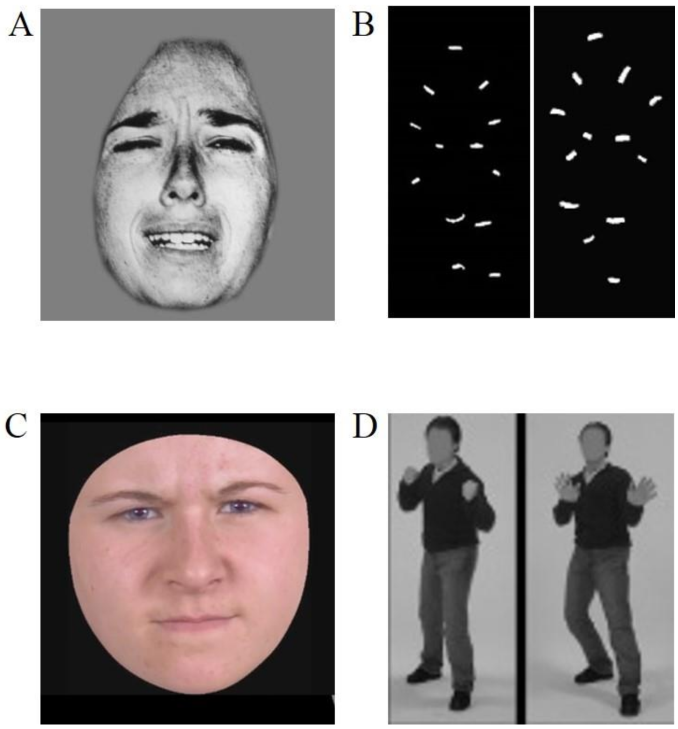
Stimuli

Examples of each of the stimulus sets are shown in Figure 1. Examples of the two dynamic stimulus sets are presented in a video file in the Supplementary Materials.

***Face stimuli (static)*:** To capture individual differences in facial expression recognition abilities we used static image stimuli taken from the Facial Expressions of Emotion: Stimuli and Tests (FEEST) set (Young, Perrett, Calder, Sprengelmeyer, & Ekman, 2002). In brief, a total of 10 identities each posing five basic emotions (anger, disgust, fear, happiness, and sadness) were selected from the Ekman and Friesen series of Pictures of Facial Affect (Ekman & Friesen, 1976). In order to avoid floor/ceiling effects, we piloted examples of each emotional expression morphed relative to the neutral expression of the same identity using Psychomorph (Tiddeman, Burt, & Perrett, 2001). This procedure is known to lead to changes in the perceived intensity (and hence recognisability) of emotion (Calder, Young, Rowland, & Perrett, 1997). Here it was used to create five intensities (25%, 50%, 75%, 100%, and 125%) of each prototype (100%) expression (total n=250 images). In a pilot experiment comprising undergraduate and postgraduate students at the University of York (n=12 participants: 4 males), we tested recognition accuracy for each of these stimuli in a five-alternative forced choice paradigm with a 1000ms exposure time. This step is of considerable importance as the limited scope of individual differences research on emotion recognition ability has meant that suitable stimuli (i.e., free of ceiling effects and with adequate variance for individual differences research) have usually been unavailable. We then selected sets of 10 stimuli for each emotion (i.e., total N=50) that showed adequate means and variances based on these pilot data (i.e., that were not showing clear floor or ceiling effects). In the cases where a surplus of stimuli was available, we chose 10 that varied in gender and age as much as possible. This selection approach was also applied to the remaining stimulus sets.

***Body Stimuli (dynamic)*:** To capture emotion recognition ability from body expressions we used dynamic point-light walker stimuli previously described by Atkinson, Dittrich, Gemmell, and Young (2004). In short, 10 actors were recorded performing each of five emotions at three levels of intensity (typical, exaggerated, very exaggerated). Actors wore suits with 13 reflective patches on key joints of their body. Subsequent rendering removed all information other than the patches from each video, resulting in a short clip of 13 light points whose combined movement simulated the natural, dynamic expression of a human emotion. Video clips lasted between 4.2 and 8 seconds. As with the face stimuli, we chose 10 stimuli for each emotion (i.e., total N=50) that showed adequate means and variances following a pilot experiment (n=12 participants: 6 males).

Figure 1: Examples of each of the stimuli used: (A) static facial expressions, (B) frames from the dynamic body point-light displays, (C) a frame from the dynamic facial expressions, and (D) static body expressions. Note: Stimulus sets A and B were used in Studies 1, 2 and 3; Stimulus sets C and D were only presented in Study 2.



Procedure

Stimuli were blocked according to expressive domain. Face and body blocks were each presented twice to the participants in a fixed order (face-body-face-body). In a five-alternative forced choice paradigm, participants had to select the emotion they thought was displayed by each stimulus using radio buttons on screen. Each face stimulus was presented for 1000ms. This limited exposure time was chosen in order to increase difficulty of the recognition task, and to ensure suitability of the data for individual differences research (i.e., by increasing error rates and variation in recognition accuracy). Body stimuli were presented for the duration of each video clip. Participants could provide their response at any point following the onset of the stimulus presentation. The within-block presentation order was fully randomised.

Participants were given the opportunity to rest following completion of each block. The mean recognition performance across blocks for each emotion was used in our analyses.

Analysis

In order to assess possible sex differences in emotion recognition from face and body, we used the data from Study 1 to conduct a three-way repeated measures Analysis of Variance (ANOVA) (2×2×5) with sex (male, female) as a between-subjects factor, and expressive domain (face, body) and emotion (anger, disgust, fear, happiness, sadness) as repeated within-subjects measures. For significant interactions, we conducted post-hoc tests to further analyse the results. As noted earlier, the prior literature did not provide strong bases for hypothesis testing and so these analyses were exploratory in nature.

6.2.2 Results

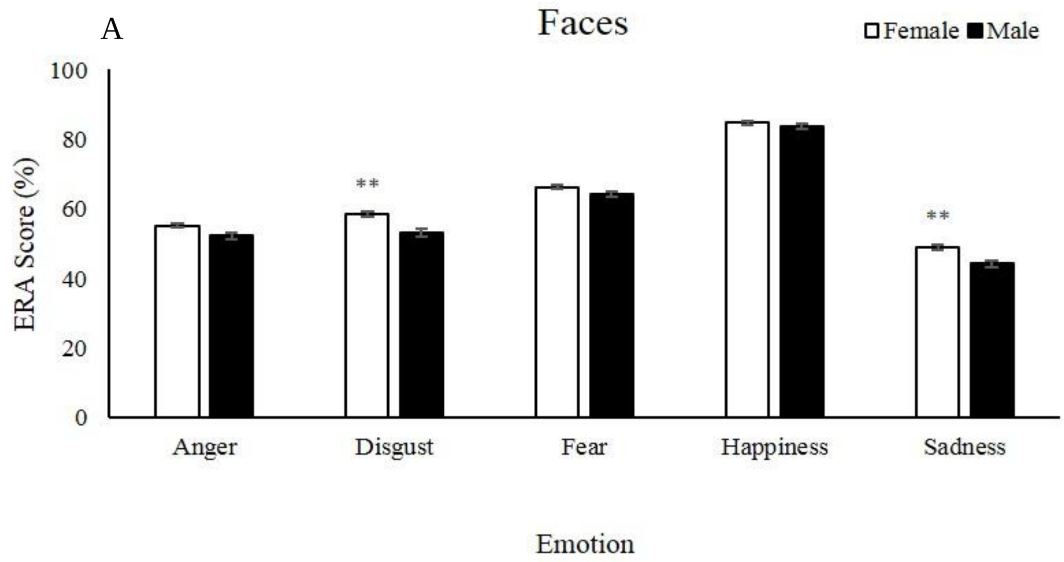
The data were submitted to a three-way ANOVA exploring the effects of sex, domain and emotion. There was a significant main effect of domain ($F(1, 1020) = 46.37$, $p<.001$, partial $\eta^2=.04$ [CI95%: .02-.07]) with emotional expressions from faces being more accurately recognised (M=.61, SD=.17) than expressions from bodies (M=.59, SD=.18). For the factor of emotion, Mauchly's test indicated that the assumption of sphericity had been violated (χ^2 (9)= 81.91, $p<.05$). Therefore, we corrected degrees of freedom using Greenhouse-Geisser estimates of sphericity ($\varepsilon=.96$). The main effect of emotion was significant ($F(3.84, 3916.41) = 600.08$, $p<.001$, partial $\eta^2=.37$ [CI95%: .35-.39]) indicating differential performance across the five emotions. Recognition accuracy was greatest for happiness (M=.69, SD=.16), followed by fear (M=.66, SD=.17), anger (M=.61, SD=.17), sadness (M=.59, SD=.18), and disgust (M=.45, SD=.19). The main effect of sex was non-significant ($F(1,1020) = 2.96$, $p=.086$, partial $\eta^2=.003$ [CI95%: .00-.01]).

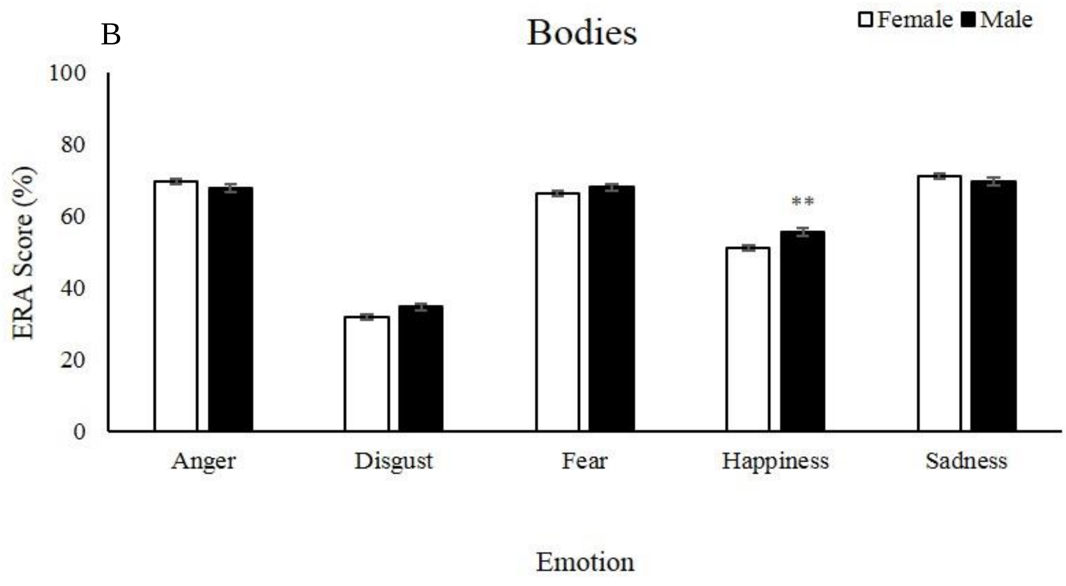
There was a significant interaction between expressive domain and sex ($F(1, 1020) = 33.32$, $p<.001$, partial $\eta^2=.03$ [CI95%: .01-.06]). Further significant interactions were also found between domain and emotion ($F(3.78, 3859.82) = 1108.06$, $p<.001$, partial $\eta^2=.52$ [CI95%: .50-.54]), and between sex and emotion ($F(3.84, 3916.41) = 6.28$, $p<.001$, partial $\eta^2=.006$ [CI95%: .002-.011]). These interactions were also corrected using Greenhouse-Geisser estimates ($\varepsilon = .95$). Of importance, these main effects and two-way interactions were qualified by a significant three-way interaction (also Greenhouse-Geisser corrected) between expressive domain, emotion, and sex ($F(3.78, 3859.82) = 3.69$, $p=.006$, partial $\eta^2=.004$ [CI95%: .003-.007]), which we now explore in greater depth.

We examined the effects of emotion and sex for each domain separately, by running mixed ANOVAs with emotion as a within-subjects factor and sex as a between-subjects factor for face and body stimuli separately. In both domains, the emotion variable violated the assumption of sphericity, therefore Greenhouse-Geisser corrections were used ($\varepsilon_{Faces}=.95$, $\varepsilon_{Bodies}=.91$). For facial stimuli, there were significant main effects of emotion ($F(3.82, 3891.79) = 762.72$, $p<.001$, partial $\eta^2=.43$ [CI95%: .41-.45]) and sex ($F(1, 1020) = 25.03$, $p<.001$, partial $\eta^2 =.02$ [CI95%: .009-.045]). These effects were qualified by a significant interaction between emotion and sex ($F(3.82, 3891.79) = 2.99$, $p=.020$, partial $\eta^2=.003$ [CI95%: .001-.006]). For body stimuli, the ANOVA revealed a significant main effect of emotion ($F(3.64, 3717.38) = 902.14$, $p<.001$, partial $\eta^2=.47$ [CI95%: .45-.49]) and a significant interaction between emotion and sex ($F(3.64, 3717.38) = 7.22$, $p<.001$, partial $\eta^2=.007$ [CI95%: .002-.013]) but no main effect of sex ($F(1, 1020) = 1.91$, $p=.167$, partial $\eta^2 =.002$ [CI95%: .00-.011]).

To probe the nature of each of the emotion and sex interactions we ran ten post hoc (Bonferroni-corrected: adjusted $\alpha = .005$) t-tests (i.e., one for each emotion across both expressive domains) comparing male and female performance (see Figure 2). These analyses revealed that females performed significantly better on recognition of facial expressions of disgust (*t*(1020) = 4.19, *p*<.001, Cohen's *d* = .28 [CI95%: .15-.41]) (female M=.58, SD=.19; male M=.53, SD=.19), and sadness (*t*(1020) = 3.90, *p*<.001, Cohen's *d* = .26 [CI95%: .13-.40]) (female M=.49, SD=.18; male M=.44, SD=.18), and that males performed significantly better for bodily expressions of happiness (*t*(1020) = 3.74, *p*<.001, Cohen's *d* = .25 [CI95%: .12-.38]) (female M=.51, SD=.18; male M=.56, SD=.17). The remaining t-tests were non-significant at our Bonferroni-corrected alpha level (all $t \leq 2.45$, all $p > .02$).

Figure 2: Mean emotion recognition accuracy in females and males in response to facial (A) and bodily (B) expressions for the five basic emotions in Study 1 (N = 1022 participants). Error bars represent standard error, and asterisks above the bars represent the results of the corrected t-tests, where ** indicates *p*<0.001.





6.2.3 Discussion

Study 1 examined the presence of sex differences in recognising emotion across face and body stimuli in a large sample. The results indicated that whilst sex had no overall effect, significant interactions between sex, expressive domain, and emotion did emerge. Upon further inspection of the nature of these interactions, we observed a significant female advantage for recognising facial disgust and facial sadness, and a significant male advantage for recognising bodily happiness. No other significant sex differences were noted.

We next sought to confirm the key findings observed from Study 1 in a replication sample. In addition, we sought to assess whether these findings extended to reflect emotion processing more broadly. To this end we used data from a sample of participants who had completed an emotion recognition battery involving the same tests reported above and two novel stimulus sets: specifically, dynamic facial stimuli and static body stimuli.

6.3 Study 2

6.3.1 Methods

Participants

In Study 2 we analysed a set of archival data from an independent MTurk sample, also drawn from the same previous study (as reported in Lewis et al., 2016: Study 3, N=384). Again, we only included participants in our analyses who completed at least 90% (≥17 of 19: note, this study did not assess static bodily disgust – see below for further details) of the trial blocks for each emotion and expressive domain, and showed no evidence of false responding, leading to the omission of 81 participants and a final sample size of 303 (137 males). The gender ratio was near-identical across White and non-White participants. In order to test the specific effects observed in Study 1, this sample provided power of .69, .72, and .78 (for a Cohen's *d* of .25, .26, & .28, respectively) for a one-tailed t-test with an alpha level of .05. Mean age was 34.8

years (SD = 11.3). A range of ethnicities were reported in the sample: White (n=232), Hispanic (n=16), Asian (n=16), Black (n=10), Native American (n=1), Other (n=15), and Undisclosed (n=13).

Stimuli

We used the same stimuli as described above, together with the additional sets of face and body stimuli detailed next. See Figure 1 for picture examples of each of our four types of stimuli, and the video in the Supplementary Materials for examples of the dynamic stimuli.

Face stimuli (dynamic): We used a sub-set of dynamic facial stimuli previously used for emotion recognition work (Lau et al., 2009). In brief, these stimuli were created by morphing one male and one female image from a neutral expression to one of the five basic emotions (anger, disgust, fear, happiness, and sadness) and then assembling the morphed images into a video clip. Stimuli dynamically changed from the neutral expression to one of four levels of intensity (25%, 50%, 75%, 100%), but for happiness, due to the ceiling effects often observed, intensity levels were lower (10%, 25%, 50%, and 75%). We only used stimuli where actors directly faced the camera, with either direct or averted gaze. This led to a total of 80 stimuli that we piloted as before, in order to avoid floor/ceiling effects (n=47 participants: 28 males, recruited through Amazon MTurk) before selecting sets of 10 stimuli for each emotion (i.e., total N=50) that showed adequate means and variances based on these pilot data. Each video clip was approximately 1.5 seconds in length.

Body stimuli (static): To test emotion recognition from static bodies we employed the Bodily Expressive Action Stimuli Test (BEAST) stimuli set (de Gelder & Van den Stock, 2011). In brief, these static stimuli comprised black and white whole body photographs of actors with faces obscured depicting one of four basic emotions (anger, fear, happiness, and sadness). Disgust is not included in this stimulus set due to it being difficult to represent in the static body alone (de Gelder & Van den Stock, 2011). The original image set contains 254 images. We again undertook undergraduate piloting at the University of York (n=14 participants: 6 males) to identify 10 stimuli per emotion (i.e., total N=40) suitable for an individual differences task, for which we then validated means and variance in a second pilot study using MTurk participants (n=50). As with the static facial images, we presented each image for 1000ms.

Procedure

The procedure was similar to that outlined for Study 1 with the exception that participants completed the following blocks in fixed order: static bodies, static faces, dynamic bodies, and dynamic faces. As such, each stimulus set was only seen once (as opposed to twice in Study 1). We describe Study 2 as a replication sample in the sense that we used the same stimulus sets in our second sample as we did in Study 1 (alongside two additional stimulus sets to test for the generalisability of the effects), as well as only examining the significant results that emerged in Study 1.

Analysis

Here we attempted to replicate only the significant effects observed in Study 1 to assess their robustness, as well as to examine whether these effects generalised to the additional stimulus sets. We chose to constrain our analyses to only the significant effects from Study 1, as any effects emerging in Study 2 that were not observed in Study 1 were unlikely to be of substantive interest given the much larger sample size of Study 1.

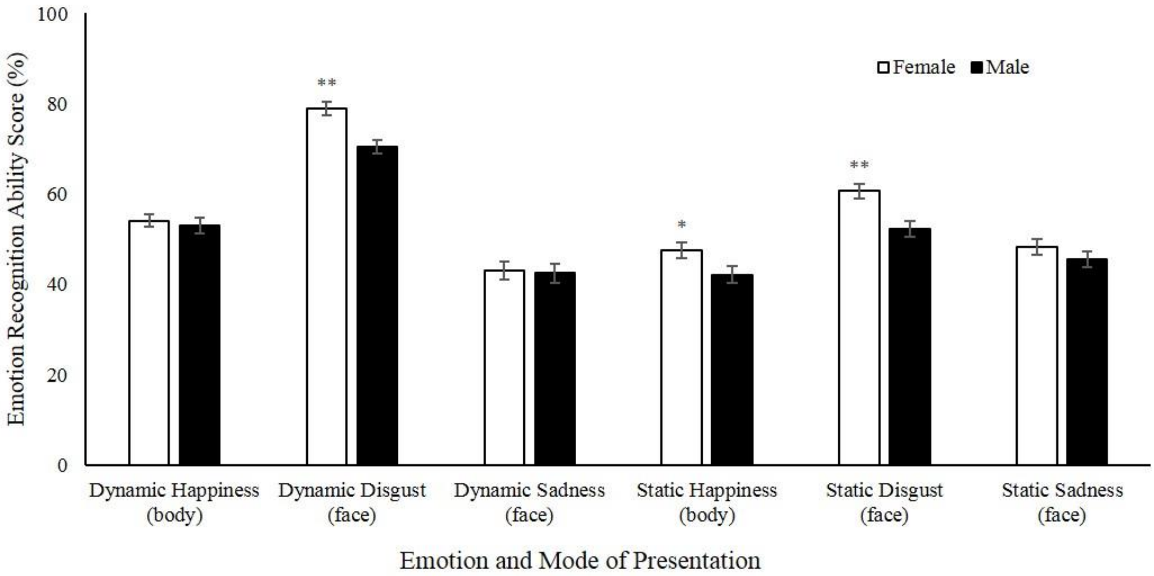
6.3.2 Results

We ran six t-tests based on the three significant findings from the discovery sample (i.e., facial disgust, facial sadness and bodily happiness) and tested each of these in both static and dynamic forms. Therefore, our six t-tests comprised static facial disgust, dynamic facial disgust, static facial sadness, dynamic facial sadness, static bodily happiness, and dynamic bodily happiness.

As in Study 1, females performed significantly better on recognition of static facial expressions of disgust ($t(301) = 3.54$, $p<.001$, Cohen's $d = .41$ [CI95%: .18-.64]) (female M=.61, SD=.21; male M=.52, SD=.21). In addition, they also performed better than males on recognition of dynamic facial expressions of disgust ($t(301) = 4.09$, $p<.001$, Cohen's $d = .47$ [CI95%: .24-.70]) (female M=.79, SD=.18; male M=.70, SD=.18).

In contrast to Study 1, females scored significantly higher on recognition of happiness from static bodies ($t(301) = 2.09$, $p=.038$, Cohen's $d = .24$ [CI95%: .01-.47]) (female M=.48, SD=.23; male M=.42, SD=.22). However, we observed no differences between males and females for dynamic expressions of bodily happiness ($t(301) = .49$, $p=.625$, Cohen's $d = .06$ [CI95%: -.17, .28]) (female M=.54, SD=.18; male M=.53, SD=.20). Finally, in contrast to Study 1 we observed no difference between males and females for either static facial expressions of sadness ($t(301) = 1.09$, $p=.275$, Cohen's $d = .13$ [CI95%: -.10, .35]) (female M=.48, SD=.23; male M=.46, SD=.21) or for dynamic facial expressions of sadness ($t(301) = .21$, $p=.835$, Cohen's $d = .02$ [CI95%: -.20, .25]) (female M=.43, SD=.27; male M=.42, SD=.25). These results are also detailed in Figure 3.

Figure 3: Mean emotion recognition accuracy in females and males in response to dynamic and static facial and bodily expressions for the three basic emotions that reached significance in Study 2 (N = 303 participants). Error bars represent standard error, and asterisks above the bars represent the results of the t-tests, where * indicates $p<0.05$ and ** indicates $p<0.001$.



6.3.3 Discussion

The results from Study 2 confirm the significant female advantage for facial disgust that was observed in Study 1. Of importance, this sex difference in disgust recognition was also observed in an additional stimulus set comprising dynamic facial stimuli. In addition, females scored significantly higher on recognition of static bodily happiness, although this should be interpreted in the context of a significant male advantage in Study 1. No other sex differences emerged.

Although these results confirm the finding of Study 1 with regards to sex differences in disgust recognition ability, our statistical power was lower than conventional levels (i.e. below 80%). This will have limited our ability to detect sex differences in facial sadness and bodily happiness, as well as raised the potential for a false positive in our disgust recognition observation. With this in mind we conducted a third study with a sample size adequately powered to reliably detect effects of the magnitude we observed in Study 1.

6.4 Study 3

6.4.1 Methods

Participants

In order to assess the robustness of the effects observed in Studies 1 and 2, we collected a third independent sample of MTurk participants. We sought to recruit a sample that would provide power of .80 to detect a Cohen's *d* of .20 in a one-tailed test with alpha at .05. This level of power was chosen with the concern that the previously observed effect sizes – particularly of Study 1 with its larger sample – might be overestimates of the population effect size. To this end we continued recruiting until we had usable data that satisfied this power requirement (note, because the required sample size for a given level of power is sensitive to the gender ratio, and because we of course could not precisely know this ratio ahead of time for our final sample, we assessed the power of our sample – with our data exclusion protocol in place – periodically throughout recruitment). As before, we only included participants in our analyses if they completed 90% (≥9 of 10) of the trial blocks for each emotion and expressive domain, and showed no evidence of false responding. In total 730 participants completed our survey. 96 participants were omitted following our data exclusion protocol. Our final sample size was thus N=634 (275 males) and this sample provided power of exactly .80 to detect a Cohen's *d* of .20 in a one-tailed test with alpha at .05. The gender ratio was near-identical across White and non-White participants. The mean age was 36.8 years (SD=10.9), and ethnicity was reported as follows: White (n=490), Hispanic (n=34), Asian (n=38), Black (n=59), Native American (n=3), Other (n=8), and Undisclosed (n=2).

Stimuli

For this sample, we used the same stimuli as detailed in Study 1; specifically there were 10 static faces and 10 dynamic bodies for each of the five basic emotions (i.e., N=50 for each expressive domain).

Procedure

The procedure for Study 3 was the same as in Study 1, with the exception that each block was presented only once to the participants (as opposed to being presented twice). Participants completed the two blocks in the same fixed order: static faces, dynamic bodies. Within-block presentation order was fully randomised. The stimuli presentation and response procedure was the same as outlined in the previous two samples.

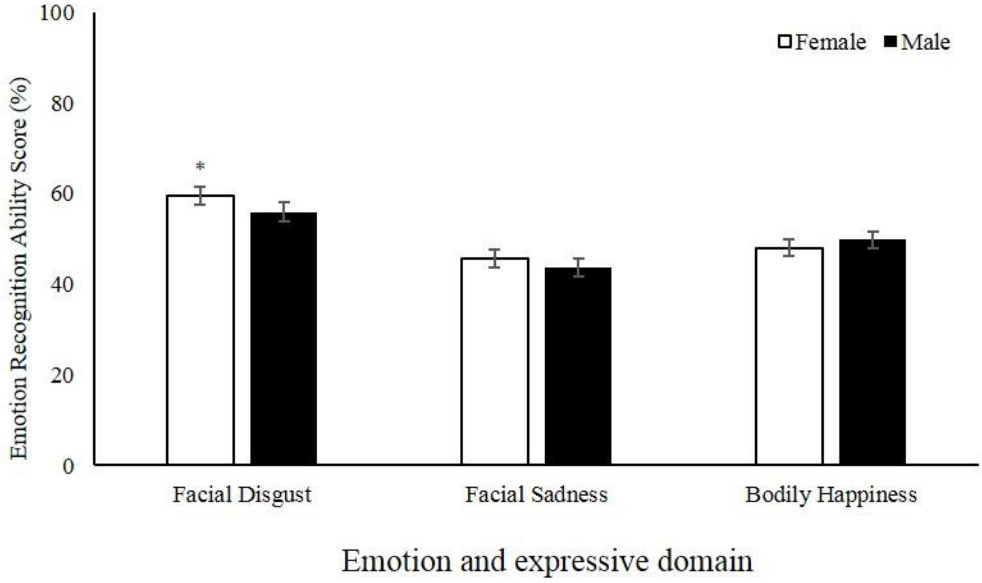
Analysis

Here we attempted to replicate only the significant effects observed in Study 1 to assess their robustness.

6.4.2 Results

We ran three t-tests based on the three significant findings that emerged in Study 1 (i.e., facial disgust, facial sadness, and bodily happiness). The tests confirmed a significant female advantage for static facial disgust ($t(630) = 2.21$, p=.027, Cohen's d = .18 [CI95%: .02-.33]) (female M=.60, SD=.20; male M=.56, SD=.21). There was no significant sex difference in recognition of facial sadness ($t(630) = 1.22$, p=.223, Cohen's d = .10 [CI95%: -.06, .26]) (female M=.46, SD=.20; male M=.44, SD=.20), or bodily happiness ($t(630) = 1.19$, p=.234, Cohen's d = .10 [CI95%: -.06, .25]) (female M=.48, SD=.18; male M=.50, SD=.18) (see Figure 4).

Figure 4: Mean emotion recognition accuracy in females and males in response to facial and bodily expressions for the three basic emotions that reached significance in Study 3 (N = 634 participants). Error bars represent standard error, and asterisks above the bars represent the results of the t-tests, where * indicates p<0.05.



6.4.3 Discussion

The results of Study 3 confirm the female advantage for recognising facial disgust observed in both Studies 1 and 2. We did not see evidence for a sex difference in facial sadness and bodily happiness.

6.5 General Discussion

The current study sought to determine the role of sex in emotion recognition ability across three large samples of adults. In Study 1 (our discovery sample), while no overall main effect of sex was observed, significant interactions across sex, expressive domain, and emotion were noted. Following decomposition of these interactions, we observed that females performed significantly better than males on recognition of facial disgust and facial sadness, and that males performed significantly better than females in recognising bodily happiness. Results from Study 2 (our first replication sample) confirmed the female advantage for facial disgust. Importantly, this significant difference was present both for static disgust stimuli (as used in Studies 1 and 2) and for dynamic facial disgust stimuli (used only in Study 2). Results from Study 3 (our second replication sample) also confirmed the female advantage for static facial disgust. This pattern of findings shows that across three independent samples, with different presentation formats (static or dynamic), females consistently outperformed males with regard to recognising facial disgust stimuli.

In contrast to this consistent female advantage for recognising facially expressed disgust, the findings in Study 1 of a female advantage for sad faces and of a male advantage for happy bodies were not replicated. In fact, in Study 2 we observed the opposite effect: females were significantly better at recognising happiness from static bodies (with no sex differences observed for dynamic happy bodies), and in Study 3, neither facial sadness nor bodily happiness showed significant sex differences. These results indicate that the findings in Study 1 probably reflect random sampling variability and so we do not discuss these failures to replicate any further.

In sum, then, these findings support the existence of a robust sex difference for facial disgust but not for any of the other basic emotions. It should be noted, however, that the effect sizes observed are not large, and thus the distributions of scores for male and female participants are largely overlapping. As such, the sex difference should not be overstated; the two sexes appear to be more similar on this ability than they are different.

Why do our findings diverge from what might be thought of as conventional wisdom, i.e., that there is an overall sex difference in emotion recognition? One possible explanation is that of publication bias in this field. This account is supported by a recent meta-analysis of sex differences in emotion recognition ability that reported evidence for an excess of significant findings in the literature (Thompson & Voyer, 2014). For the field to move towards a consensus state, this suggests a need for strongly powered confirmatory studies with pre-registered experimental protocols.

What might account for this modest female advantage for recognising facial disgust? An interesting, if at present speculative, perspective posited to account for this more general sex difference concerns the unique selection pressures faced by females, including immunosuppression during pregnancy and over the menstrual cycle, higher risk of contracting sexually-transmitted diseases and transferring them to their offspring, and higher parental investment in infant protection (Fleischman, 2014). Given the greater vulnerability of females and their altricial offspring to contamination and infection, an evolutionary functionalist account theorises that females need to be more sensitive to cues of disgust (Curtis, Aunger, & Rabie, 2004), which may include facial expressions.

In addition, Fleischman (2014) also hypothesised that males may show less disgust sensitivity than females due to a selection pressure to emphasise their robust immunity by displaying a relative indifference towards signals of disgust. Given that males consistently show greater risk-taking behaviour than females, Fessler, Pillsworth and Flamson (2004) suggested that this propensity may also extend to a higher willingness to approach sources of contamination in comparison to their female counterparts. By employing the minimum possible level of disease avoidance, it is suggested that males are highlighting their successful immune system and high genetic quality to potential reproductive mates.

It is conceivable, then, that such fitness imperatives shape the ability to recognise disgust in conspecifics in different ways across the sexes. The evolutionary adaptationist theory that disgust sensitivity may be functionally related to successful mating and reproduction is supported by the finding that sex differences are observed from puberty and young adulthood onwards, but no sex differences in disgust sensitivity emerge in child participants (Stevenson, Oaten, Case, Repacholi, & Wagland, 2010), as well as the finding that disgust sensitivity decreases across the lifespan as reproductive potential declines (Curtis et al., 2004). That said, we must reiterate that our study found largely overlapping distributions of ability to recognise disgust across women and men, so any evolutionary influences do not create major differences in this respect. Moreover, it is easy to think of ways in which cultural socialisation might tend to make women more sensitive than men to the importance of hygiene and risk of disease.

Some limitations of our studies require mention. Firstly, while MTurk samples are more diverse than student samples (Paolacci et al., 2010), they clearly do not form a representative sample of the population; for example, there was a greater proportion of female than male respondents in all three of our studies. And although our samples were ethnically diverse, they were mainly White and comprised solely of US residents; as such, all participants have been strongly exposed to Western culture. Further studies involving non-Western populations will therefore also be valuable in order to assess the robustness of the sex differences reported here. Secondly, here we focused on five basic emotions; however, emotional expressions are of course not restricted to these categories. Accordingly, work that explores a broader selection of emotions may reveal additional sex differences. Thirdly, while our stimuli are among the most carefully developed and validated for laboratory-oriented research, they clearly have limits with regards to ecological validity. As such, work that can take advantage of more naturalistic stimuli will be of value in future studies. Finally, we note that our bodily disgust recognition rate was low (in fact close to chance level) for both sexes, and this high level of noise may have led to a false negative result. Future work is recommended to address this issue.

Conclusions

In summary, across three independent samples we observed that females are superior to males in facial disgust recognition. This result is of particular note as this sex difference was observed across two very different sets of facial stimuli (static and dynamic). However, these group differences were modest in magnitude, and the overlap in ability between the two populations is substantial. No consistent evidence for further sex differences in emotion recognition ability was observed.

Chapter 7: Emotion Recognition across Adult Age

Face perception across the adult lifespan:

Evidence for age-related changes independent of general intelligence

The work presented in this section was conducted in collaboration with Andrew Young, and Gary Lewis (supervisor) and is under review at *Cognition and Emotion*.

Abstract

It is well-documented that face perception – including expression and identity recognition ability – declines with age. To date, however, it is not yet well understood whether this age-related decline reflects face-specific effects, or instead can be accounted for by well-known declines in general intelligence. We examined this issue using a relatively large, healthy, age-diverse (18-88 years) sample (N=595) who were assessed on well-established measures of face perception and general intelligence. Replicating previous work, we observed that facial expression recognition, facial identity recognition, and general intelligence all showed declines with age. Of importance, the age-related decline of expression and identity recognition was present even when the effects of general intelligence were statistically controlled. Moreover, facial expression and identity ability each showed significant unique associations with age. These results indicate that face perception ability becomes poorer as we age, and that this decline is to some extent relatively focal in nature. Results are in line with a hierarchical structure of face perception ability, and suggest that age appears to have independent effects on the general and specific face processing levels within this structure.

7.1 Introduction

Face perception encompasses a range of abilities that are necessary for successful everyday interactions (Bruce & Young, 2012). Among these abilities, the perception of expression and identity are of critical importance. Accurate perception of facial expressions is essential for appropriate responses to the subtle and rapid changes in a person's demeanour and emotional state, whilst accurate identification of others via their face enables social interactions to be appropriately adjusted based on prior knowledge of and previous encounters with that individual (Young, 2018).

Of note, then, a substantial body of research has consistently reported a decline in face perception abilities with age. In the context of facial expression recognition, a meta-analysis of 28 datasets (total N=1667) reported age-related decline in face emotion recognition that was evident across categories of emotions (Ruffman, Henry, Livingstone, & Philips, 2008). In a sample of 607 participants (18-84 years) who were tested on facial and vocal expression recognition, older participants were shown to be less accurate across emotions (Mill, Allik, Realo, & Valk, 2009). In a sample of 482 participants (20-89 years), participants in their 30's, 40's and 50's showed equivalent accuracy in expression recognition, but a linear decline was seen to emerge from 60 years of age onwards, and further declines were particularly noticeable for participants in their 70's and 80's (West et al., 2012). In a large study (N=7230, 18-75 years), Sasson and colleagues observed a deficit for older adults' expression recognition across all tested emotions (Sasson et al, 2010). Another sample (N=9546, 10-85 years) observed age-related deficits in emotion sensitivity (i.e. discriminating between the intensity of two expressions) (Rutter et al., 2019). Finally, a very large community sample (N=100,257) reported age-related deficits on an emotion recognition task involving composite expressions in a sample of individuals who ranged from younger than 15 years of age to older than 60 years, with the older groups performing worse than their younger counterparts (Olderbak, Wilhelm, Hildebrandt, & Quoidbach, 2018).

In the context of facial identity recognition, age-related changes have also been noted. In a sample of 448 participants (18-88 years), Hildebrandt and colleagues (2010) observed considerable age-related performance decrements across three aspects of identity recognition: face memory (e.g. immediate and delayed recognition of learned faces), face perception (e.g. part-whole matching tasks), and speed of face identity matching (e.g. matching of faces from different viewpoints). Decrements were strongest for the speed of face identity matching (showing a linear decrease beginning in the early 30's) but were also apparent for memory (the late 40's) and perception (the 60's). Age-related decrements have also been reported for another unfamiliar face matching task (Benton, Eslinger, & Damasio, 1981), and in holistic perception (Boutet & Faubert, 2006). In eyewitness identification paradigms, older adults show lower accuracy on line up tasks, and a higher rate of false recognition of new faces (Searcy, Bartlett, & Memon, 1999).

This body of work provides strong evidence for an age-related decline in face-related abilities. However, it is not yet known if this decline reflects changes in face perception per se, or instead is simply a reflection of well-known age-related declines in general intelligence (Deary, 2001; Salthouse, 2010). A huge amount of empirical research demonstrates the significant age-related declines observed in the domains of reasoning, spatial visualisation, verbal memory and perceptual speed, with vocabulary in contrast showing an increase or preservation until approximately age 60 (Salthouse, 2013). The possibility that this general cognitive decline underpins age-related decline in face perception abilities is bolstered by evidence from recent research demonstrating robust links from general intelligence to both expression recognition (Hildebrandt, Sommer, Schacht, & Wilhelm, 2015; Lewis, Lefevre, &

Young, 2016; Schlegel & Scherer, 2016) and identity recognition (Wilhelm et al., 2010; Shakeshaft & Plomin, 2015; Connolly, Young, & Lewis, 2019).

A handful of studies have already attempted to address this issue, although typically without a direct measure of general intelligence. In the context of expression recognition, Mill and colleagues (2009) observed that age remained a significant predictor when adjusting for education, a proxy for general intelligence (Deary, Strand, Smith, & Fernandes, 2007). West and colleagues (2012) reported the age/expression recognition association even when adjusting for processing speed, which is moderately associated with general intelligence (Neisser et al., 1996; Sheppard & Vernon, 2008). Horning, Cornwell, and Davis (2012) used the Raven's matrices reasoning test as a proxy for fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999), and found that whilst this was a significant predictor of recognition of some of the basic emotions, age also remained a significant predictor. Finally, in terms of identity recognition, Hildebrandt and colleagues (2011) reported that age-related differences in memory for faces were still evident after adjusting for age-related differences in general cognition, as measured by Raven's advanced progressive matrices, and two working memory tasks: a rotation span task and a memory updating task.

However, as alluded to above, a crucial caveat to these studies is the measure of intelligence. In most of the larger studies only a proxy for general intelligence has been used, such as years or level of education (Mill et al., 2009; Sasson et al., 2010; Kessels et al., 2014), matrix reasoning (Horning et al., 2012), or processing speed (West et al., 2012). Whilst these variables undoubtedly correlate with general intelligence (e.g. Deary, Der, & Ford, 2001), it is important to note that they fail to fully capture the broad variance of this ability. It is plausible, then, that if a more comprehensive measure was included, it might completely attenuate the association between age and face perception.

The Current Study

The current study sought to offer clarity regarding this important issue by leveraging data from a relatively large age-diverse sample who had been assessed on well-acknowledged measures of face perception and general intelligence. The Cattell Culture-Fair Intelligence Test comprises four nonverbal subtests of matrix reasoning, and whilst the constructs of fluid and general intelligence have been debated in the field, factor analytic research has shown very strong correlations (r = .77-96) between the Cattell test and other more broadly constructed cognitive batteries, e.g. the General Aptitude Test Battery (Johnson, te Nijenhuis, & Bouchard, 2008), indicating a high level of common measurement across these various cognitive batteries. In comparison to single measures of matrix reasoning or processing speed, then, the Cattell test battery better captures the breadth of general intelligence, and is well suited for our specific research question regarding age-related decline.

The face perception measures included a test of emotion recognition involving morphed images to create differing levels of task difficulty (Young et al., 1997; 2002), thus making it sensitive enough to generate a range of scores and thus suitable for individual differences research in our sample of healthy adults. The measures also included the Benton Test of Facial Recognition (Levin, Hamsher, & Benton, 1975). Whilst the Benton test is based entirely on unfamiliar face recognition (Young & Burton, 2018) and there has been debate about the circumstances in which it is useful (Duchaine & Weidenfeld, 2003; Rossion, 2018) it has the advantages of being a widely-used and purely perceptual measure that generates a range of individual differences in performance. Importantly, in light of the fact that we had access to measures of facial expression and identity recognition in the same sample, the current study was able to examine whether these age-related declines showed *unique* associations with age; that is, whether face perception abilities showed a general decline with age, or whether this decline was specific to expression or identity recognition ability.

7.2 Methods

Participants

The data analysed in this study were collected by the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) (Shafto et al., 2014). The Cam-CAN sample is cross-sectional and age-diverse (aged 18-88 years). Participants completed demographic questionnaires and general cognitive and memory assessments in a home interview. Following an initial assessment, 700 eligible individuals (50 men and 50 women for every age decile) who were MRI-suitable were invited to complete a range of neuroimaging sessions and cognitive-behavioural tasks, including the cognitive measures examined in the current study. Exclusion criteria for non-eligible participants included: low cognitive health (Mini Mental State Exam score of 24 or lower); poor hearing (failing to hear 35dB at 1000 Hz in either ear); poor vision (below 20/50 on Snellen test); low English language ability (non-native or non-bilingual English speakers); self-reported substance abuse; and serious health conditions that would affect participation (for example, major neurological or psychiatric conditions, current chemo/radiotherapy, or a history of stroke. A total of 656 (291 men) participants were thus recruited and these data form the basis for the analyses reported here.

Participants were next excluded if they showed chance levels of performance on two or more of the cognitive-behavioural tasks, or had not completed all of the cognitive-behavioural tests (see Measures). Participants were also excluded if they were missing age information. This necessitated the exclusion of 61 participants, resulting in a final sample size of 595 (291 men). The mean age of participants was 54.0 years (SD=18.2, range=18-88), and ethnicity was as follows: White (N=573), Asian (N=7), Black (N=1), Mixed Race (N=8) and undisclosed (N=6).

Measures

Facial expression recognition ability was assessed using the Emotion Hexagon test (Young et al., 1997, 2002). This test was created by using a model from the Ekman and Friesen (1976) ‘Pictures of facial affect’ series displaying each of the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise). These images were then morphed with another basic emotion to form emotional expressions with graded levels of difficulty (expression pairs morphed together consist of happiness-surprise, surprise-fear, fear-sadness, sadness-disgust, disgust-anger, and anger-happiness). Participants were shown faces with either 70% or 90% of the target emotion, and had to make a six-alternative forced-choice response to indicate whether the expression was most like anger, disgust, fear, happiness, sadness, or surprise. There were 20 trials for each of the six emotions, and stimuli were shown for 3 seconds each. A percentage accuracy score for each of the six emotions was generated for use in subsequent analyses. The six Emotion Expression Recognition sub-scores were significantly associated: r ranged from .12 to .46, and all $p < .003$.

Facial recognition ability was assessed using the short-form of the Benton Test of Facial Recognition (Levin et al., 1975), which measures the ability to match pictures of unfamiliar faces. The test consists of 27 trials in which the participant is shown one target face and an array of six faces. The participant has to identify one or more examples of the target face in the array. There may be changes in head orientation or lighting between the target and array faces. Each correct response receives a score of 1, and a total percentage accuracy score was generated for use in subsequent analyses.

General intelligence was assessed using the Cattell Culture Fair Intelligence test (Scale 2 Form A: Cattell, 1973), which contains four nonverbal subtests: Series Completion, Classification, Matrices, and Conditions. Participants are given 3, 4, 3, and 2.5 minutes, respectively to complete each subtest. The test uses a pen-and-paper approach: the participant is asked to choose a response for each item from multiple response options and to record their response on a corresponding answer sheet. Correct responses are given a score of 1 and the percentage correct for each sub-test was calculated for use in subsequent analyses. The four Cattell Culture Fair Intelligence subtests were significantly associated: r ranged from .52 to .63 (all $p < .001$).

Procedure

Eligible participants attended testing sessions at the Medical Research Council Cognition and Brain Sciences Unit in Cambridge UK. Approximate duration for each of the tasks was as follows: Emotional expression recognition: 20 minutes; Unfamiliar facial recognition: 10 minutes; and General intelligence: 20 minutes. The emotional expression recognition test was presented on a laptop, and the unfamiliar facial recognition and intelligence tests were administered using pen and paper. The majority of participants were comfortable using the laptop for the emotional expression task, but if a participant struggled, the researcher pressed the buttons for them in response to their spoken answer. This ensured that the accuracy of a participant's answer would not be confounded by their computer competency.

Analysis

Measurement Invariance

As a sensitivity check, we tested for measurement invariance separately for the two variables with sufficient number of manifest variables to stably identify a latent factor (General Intelligence: four Cattell subtests; Face Expression recognition: six emotion categories).

General Intelligence

Firstly, we tested for configural invariance by examining whether the same pattern of freed and fixed parameters held across three defined age groups (Younger Adults: 18-39 (N=153); Middle-aged Adults: 40-64 (N=243); Older Adults: 65+ (N=199). Model results demonstrated that configural invariance was evident across the three age groups, (X2 (6) = 5.36, p=.499; CFI=1.00; RMSEA=.00). Secondly, metric invariance (i.e., weak factorial invariance) was tested by examining if the factor loadings were equivalent across groups. Model results in aggregate demonstrated evidence for complete metric invariance (X2 (12) = 21.64, p=.042; CFI=.98; RMSEA=.06). However, the chi square difference between this model and the configural model was significant, X2 difference (6) = 16.28, p=.012, suggesting that the metric model had a significantly worse fit.

We thus explored whether metric model fit could be improved by adjusting some model parameters. For this, we inspected the modification indices and in turn allowed the loading of the second Cattell subtest (Classification) to vary across age groups, with the other three subtest loadings remaining constrained to equality across the age groups. We re-ran the metric invariance test with this modification and found this model had excellent fit (X2 (10) = 10.84, p=.370; CFI=.99; RMSEA=.02). Moreover, the chi square difference between this partial metric model and the configural model was non-significant, X2 difference (4) = 5.48, p=.241.

The results of this invariance testing suggest that the factor structure of general intelligence is equivalent across age groups. However, complete metric invariance was not able to be established, suggesting that at least some of the age group differences in general intelligence reflect variance beyond the general factor level of analysis.

Face Expression Recognition

Secondly, we assessed whether face expression recognition ability was invariant across age. Firstly, configural invariance was established across the three age groups, X2 (21) = 29.44, p=.104; CFI=.98; RMSEA=.05).

Complete Metric (weak factorial) Invariance Testing: Face Expression Recognition

Model results in aggregate demonstrated evidence for complete metric invariance (X2 (31) = 51.95, p=.011; CFI=.945; RMSEA=.058). However, the chi square difference between this model and the configural model was significant, X2 difference (10) = 22.51, p=.013, suggesting that the metric model had a significantly worse fit.

Partial Metric (weak factorial) Invariance Testing: Face Expression Recognition

As before, we explored whether the metric model fit could be improved by adjusting some model parameters. For this, we inspected the modification indices and allowed the loading of the Happiness manifest variable to vary, with the other five emotion category loadings remaining equivalent across the age groups. We re-ran the metric invariance test with this partial constraint, and found this model did not have an acceptable fit according to the chi square statistic, X2 (29) = 44.78, p=.031, but had acceptable alternative fit indices (CFI=.958; RMSEA=.052). Moreover, the chi square difference between this partial metric model and the configural model was non-significant, X2 difference (8) = 15.33, p=.053, although interpreting p values close to the nominal threshold should be done with caution.

The results of this invariance testing suggest some relatively modest evidence of metric variance of expression recognition across age groups. Complete weak invariance was not able to be established, suggesting that at least some of the age group difference in expression recognition factor loadings is attributable to measurement bias. However, scale invariance was established for both general intelligence and face expression recognition, so we elected to continue using these measures to assess age differences as per our analysis plan.

7.3 Results

Descriptive statistics are detailed in Table 1. Inter-correlations between study variables are detailed in Table 2. Facial expression recognition showed strong positive correlations with general intelligence and facial identity recognition. Age was negatively associated with expression and identity recognition, and with general intelligence. These age relationships are also illustrated in Figure 1.

Table 1. Descriptive statistics of facial expression recognition ability (Face Exp), facial identity recognition ability (Face ID), age, and general intelligence.

	Mean	SD	Min	Max	Skew	Kurtosis
Age	54.01	18.16	18	88	-.04	-1.12
Face Exp	87.43	9.83	26.09	95.65	-1.36	1.75
Face ID	85.07	8.50	49.17	100.00	-.36	-.50
General Intelligence	69.58	14.48	62.96	100.00	-.56	-.14

Table 2. Zero-order correlations between measures of facial expression recognition ability (Face Exp), facial identity recognition ability (Face ID), age, and general intelligence. Values in parentheses reflect correlations when adjusting for general intelligence.

	Face ID	General Intelligence	Age
Face Exp	.39 (.22)	.52	-.44 (-.15)
Face ID		.42	-.46 (-.27)
General Intelligence			-.66

Note. All p <.001.

Regression Analysis

The regression analyses then enabled us to test our research question of whether the age-related decline in expression recognition or identity recognition was independent from the decline observed in general intelligence.

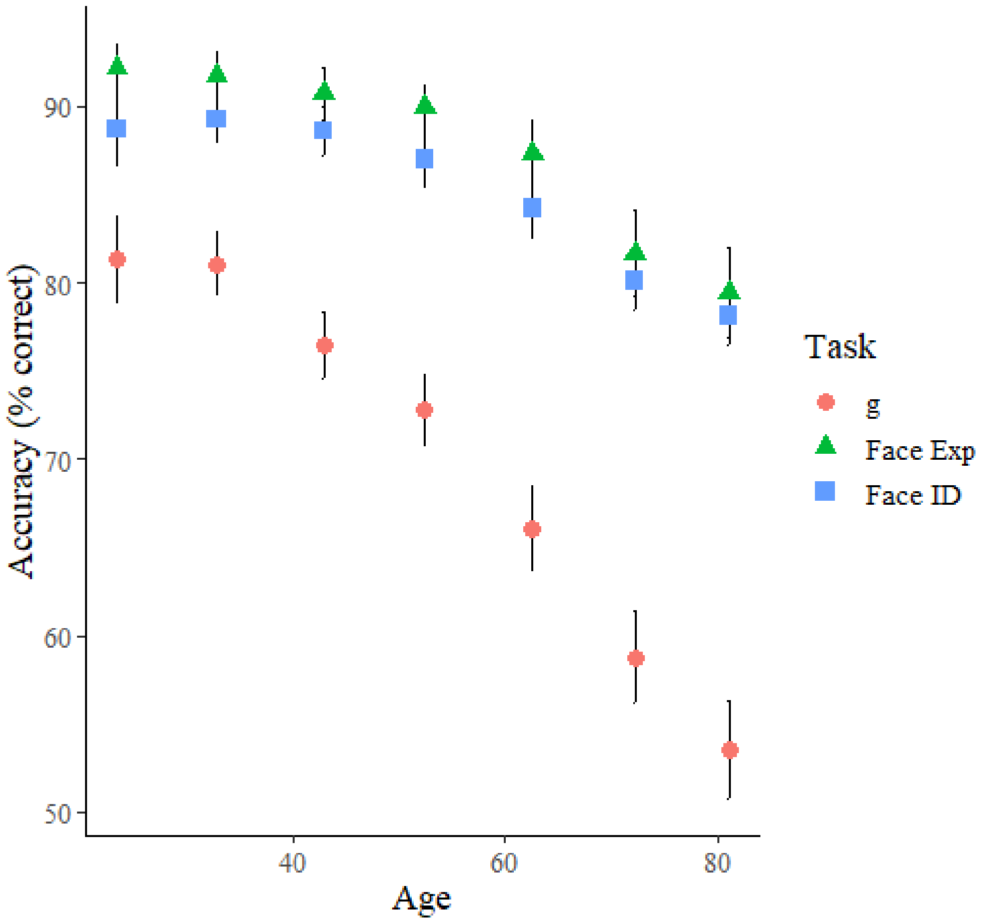


Figure 1: Relationships with age (by decile) for general intelligence (g), facial expression recognition ability (Face Exp), and facial identity recognition ability (Face ID).

In the first regression model, expression recognition was entered as the dependent variable, and general intelligence, age, sex, and identity recognition were all entered as predictors in the same step. The coefficients indicated that each of these variables was a significant and unique predictor of expression recognition. The full results of this regression analysis are shown in Table 3. Note that the coefficient for sex reflects face expression recognition scores being significantly higher for women. This finding is further analysed and discussed below.

Table 3: Multiple regressions predicting facial expression recognition ability from age, sex, facial identity recognition ability (Face ID), and general intelligence.

Independent Variables	**β**	**Sig**
Age	-.10	.04
Sex	.14	<.001
Face ID	.18	<.001
General Intelligence	.39	<.001

In the second regression model, identity recognition was the dependent variable, and general intelligence, age, sex, and expression recognition were all entered as predictors. In this case, the coefficients for general intelligence, age, and expression recognition all suggested unique influence of these variables on identity recognition, but sex was not a significant predictor. The full results of this analysis are shown in Table 4.

Table 4: Multiple regressions predicting facial identity recognition ability from age, sex, facial expression recognition ability (Face Exp), and general intelligence.

Independent Variables	**β**	**Sig**
Age	-.29	<.001
Sex	.02	.63
Face Exp	.20	<.001
General Intelligence	.13	.01

The age declines in face perception abilities when adjusting for general intelligence are further illustrated through plotting of the residuals, and are shown in Figures 2 (expression recognition) and 3 (identity recognition).

Figure 2: Relationships of age with face expression (Face Exp) recognition residuals, showing the age decline of Face Exp when adjusting for general intelligence.

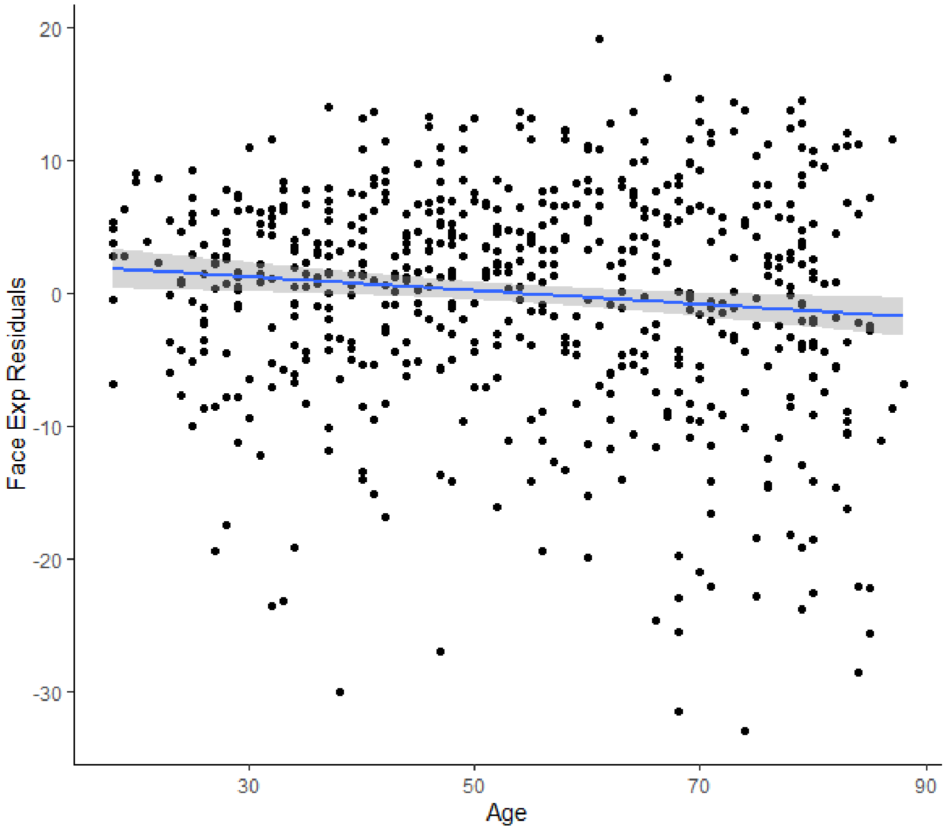
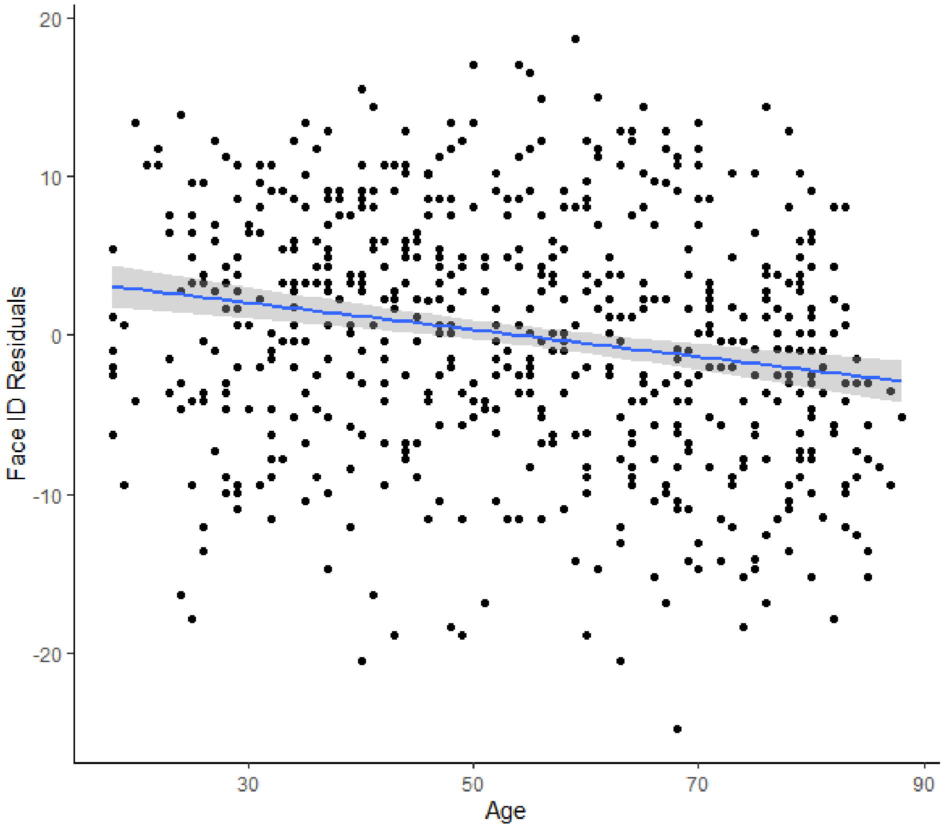


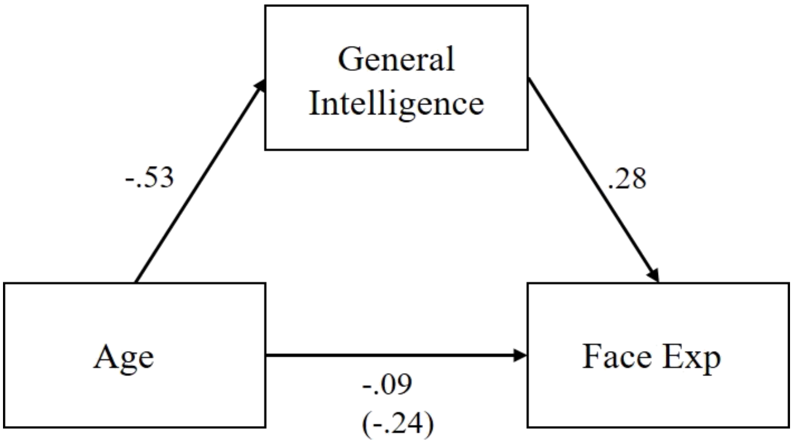
Figure 3: Relationships of age with face identity (Face ID) recognition residuals, showing the age decline of Face ID when adjusting for general intelligence.



Mediation Analyses

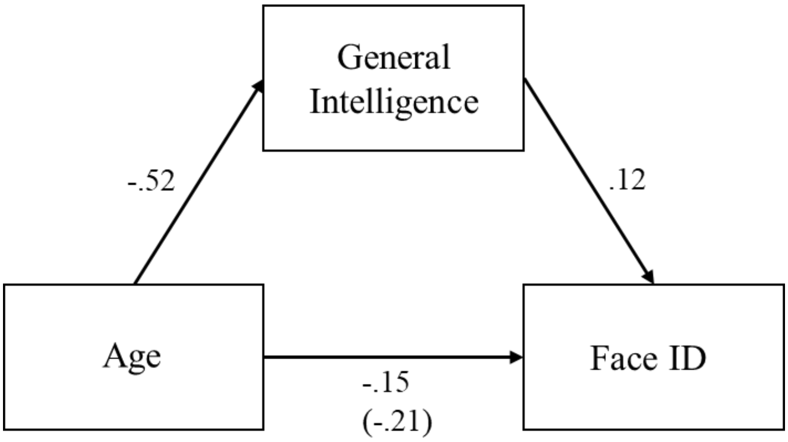
We note from the linear regressions presented above that both age and general intelligence are significant unique predictors of facial expression recognition and facial identity recognition ability. We formally tested for mediation effects using a path analysis approach implemented in the R package ‘lavaan’ (Rosseel, 2012). We tested one plausible model, whereby age was the independent variable, general intelligence the mediating variable, and facial expression or facial identity recognition as the respective dependent variable. While this arguably reflects the most theoretically plausible causal model, other possible pathway models exist, and as such we advise caution when interpreting these paths. The mediated relationships are presented in Figures 4 and 5.

Figure 4: Mediation model of age, general intelligence, and facial expression recognition ability.



Note: All standardised coefficients are significant at *p*<.001. The value in parentheses is the relationship between age and facial expression recognition before general intelligence was taken into account.

Figure 5: Mediation model of age, general intelligence, and facial identity recognition ability.



Note: All standardised coefficients are significant at *p*<.001. The value in parentheses is the relationship between age and facial identity recognition before general intelligence was taken into account.

Across the two models, age significantly predicted facial expression (β=-.09, CI=[-.14, -.04], *p*<.001), and facial identity recognition (β=-.15, CI=[-.19, -.11], *p*<.001), even when general intelligence was included in the model. The indirect effect of age through general intelligence was a significant predictor of facial expression (β=-.15, CI=[-.18, -.11], *p*<.001), and facial identity recognition (β=-.06, CI=[-.09, -.03], p<.001), indicating that general intelligence was a partial mediator of the effect between age and facial expression, and between age and facial identity recognition.

Subsidiary Analyses

Firstly, we ran some exploratory tests to examine whether age-related decline in emotion expression recognition ability instead reflected worse performance on one or more particular emotion categories. We observed moderate-to-large negative correlations of age with fear (r=-.48, p<.001), anger (r=-.31, p<.001), sadness (r=-.36, p<.001), and surprise (r=-.30, p<.001). We observed non-significant correlations of age with happiness (r=-.08, p=.042) and disgust (r=-.04, p=.326). These results suggest that the age-related decline of emotion recognition is not being driven by performance on a limited sub-set of emotions, with medium-to-large decrements observed across the majority of negative emotion categories. The non-significant correlation of age with happiness recognition is also not altogether surprising given the pervasive ceiling effects observed here (and in many other studies) that may obscure any true effects. Finally, the absence of a significant decline in disgust recognition ability across age supports the relative preservation noted in earlier work (e.g. Calder et al., 2003).

Secondly, whilst not of primary importance to the present study, the observation of a significant sex difference in favour of women for emotion recognition ability was deemed sufficiently important for us to present here for issues of replication, and thus was further explored in a subsidiary analysis. We tested whether this effect was evident across all of the emotions, or if it reflected specific emotional categories, in light of recent work noting a selective female advantage for recognising facial disgust (Connolly, Lefevre, Young, & Lewis, 2019). We subjected the data to six t-tests, correcting for multiple comparisons (Bonferroni-corrected: adjusted α=.0083). There was a significant difference in favour of women on recognition of disgust ($t(593)=-3.22$, $p=.001$, Cohen's $d=.26$, [CI95%:.10-.43]) (female M=88.59, SD=15.98; male M=83.83, SD=19.94) and in recognition of happiness ($t(593)=-3.39$, $p=.001$, Cohen's $d=.28$, [CI95%: .12-.44]) (female M=98.17, SD=4.17; male M=96.68, SD=6.37).

7.4 Discussion

A number of studies have reported age-related declines in facial expression and identity recognition abilities. However, to date, it has not been well understood if these declines reflect independent expression and identity effects, a more general face-specific effect, or simply the manifestation of the well-acknowledged decline in general intelligence observed across the lifespan (Deary, 2001). Moreover, as we have noted in the Introduction, studies to date have not used sufficiently broad measures of intelligence to fully answer this question. To address both the theoretical question and this methodological issue here, we used a relatively large cross-sectional sample of individuals ranging from 18 to 88 years of age who were assessed on widely used tests of face perception ability, and most importantly for current purposes, a comprehensive measure of general intelligence.

Expression recognition ability, identity recognition ability, and general intelligence were all negatively related to age, such that older individuals scored more poorly. Of importance, age was a significant predictor of both expression and identity recognition ability, even when general intelligence was statistically controlled, indicating that these age-related declines are not fully accounted for by the known decline in general intelligence. Indeed, the mediation analyses indicated that general intelligence was a partial mediator of the effect of age on facial expression/identity recognition. These findings are consistent with previous work that found age remained a significant predictor after adjusting for proxies of general intelligence (Horning et al., 2012; West et al, 2012). However, given that the current study used a comprehensive measure of general intelligence as opposed to a proxy measure, our results here support this finding in stronger and more concrete terms.

Whilst it is clear that different measures of face perception will often correlate with each other, such correlations are typically able to account for a maximum of around 25% of the variance across face tasks (Connolly et al., 2019; McCaffery et al., 2018; Verhallen et al., 2017). Consistent with this, the intercorrelation between the face perception variables tested here (0.39) accounted for some 15% of variance and each measure showed significant unique associations with age. This suggests that whilst the face variables themselves covary in a way that is consistent with the idea of a general factor underlying different aspects of face perception (Verhallen et al., 2017), their respective declines are to some extent independent of one another, and not solely attributable to a general overall decline in face processing ability.

This finding of independent associations of expression recognition and identity recognition to age is consistent with previous related work suggesting multiple levels of individual differences underlying face perception ability, including general intelligence, general face-specific processing, and expression- and identity-specific processes (Lewis et al., 2016; Connolly et al., 2019). This suggests then, that the effect of age may operate differently at these various levels of processing, and that at least some of the age-related decline acts upon the expression- and identity-specific level, resulting in the independent age associations that we have observed in the current study.

Our findings are also consistent with the age-related decline often observed in broader domains such as social cognition. For example, Maylor and colleagues (2002) reported a significant negative association between theory of mind (ToM) and age, when adjusting for vocabulary, processing speed, and executive functioning. More recently, Baksh and colleagues (2018) developed a test of social cognition that assessed both cognitive and affective ToM and inter- and intrapersonal understanding of social norms, and found that whilst scores on this test declined with increasing age, they were not correlated with either verbal or reasoning ability. These results suggest, then, that in line with our current face perception results, social cognition may also show a somewhat independent age-related decline from general intelligence. Further studies will be able to offer further insight into the extent of this putative independence.

In our subsidiary analysis, we noted a significant female advantage for recognising facial disgust and happiness, suggesting that the overall emotion recognition difference was being primarily driven by women's more accurate recognition of these two discrete emotions. The sex difference in disgust recognition is consistent with a recent study involving three independent samples (Connolly et al., 2019). In contrast, in that set of studies, we did not observe a female advantage for happiness recognition. This discrepancy may be the result of the more age-diverse sample reported here. It appears plausible that the sex difference in happiness recognition is stronger in older individuals and thus why it emerged in the current sample, and not in the younger adult samples tested in Connolly et al. (2019).

Some limitations of the current study are worth noting. Firstly, the design is cross-sectional, with participants being tested at only one time point, and therefore cohort effects between different generations may be a source of bias. The environments in which our younger and older participants developed are likely to vary greatly, with large differences in cultural norms and quality and quantity of healthcare, nutrition, and education, amongst other variables. Indeed, in intelligence research, the phenomenon of cognitive test scores increasing across generations has been widely established (Flynn, 1987). However, it has also been noted that within-cohort variation can be almost as large as that between different cohorts (Salthouse, 2014a), suggesting that age-related differences in cognition cannot be wholly accounted for by cohort differences. Additionally, whilst longitudinal designs have often reported positive test effects whereby participants show benefits of having had prior test experience, quasi-longitudinal designs have reported almost identical age trends to cross-sectional studies

(Salthouse, 2014b). This suggests, then, that longitudinal designs may underestimate the negative age-related change in cognition, and that cross-sectional results may be closer to estimating the true magnitude of age-related decline. Given this, then, future studies may benefit from employing longitudinal designs to corroborate the findings of cross-sectional studies.

Secondly, we must consider the likely bias induced by self-selection. The individuals comprising the current sample were recruited as part of a larger study in which participants had to attend multiple testing sessions involving MRI (magnetic resonance imaging) and MEG (magnetoencephalography) measures. Being willing and physically able to attend these sessions and complete a variety of cognitive and neural tasks suggests a certain level of motivation. In addition, in order to be eligible to take part in the neuroimaging stage of the study, the individuals had to be healthy with no serious cognitive impairment, psychiatric disorders, difficulties with vision or hearing, or evidence of substance abuse. Given the extensive cognitive and physical screening of our participants before testing (Shafto et al., 2014), the observed age-related declines in face perception are unlikely to be due to comprehension or sensory difficulties. Therefore, selected participants are likely to represent the higher end of the typical continuum in the general population, and this may be especially true for the older participants. It should be noted, though, that this source of bias would likely have led to an underestimate of the age effects observed here.

Finally, we established configural invariance of general intelligence and expression recognition, but only established partial metric invariance for these two variables. This suggests only partially equivalent factor loadings across the age groups, and that at least some of the age effect observed for intelligence and expression recognition may be accounted for by measurement bias. We suggest that future studies of expression recognition and general intelligence should seek to establish complete metric invariance of their chosen measures to ensure accurate interpretation of any age effects observed.

Conclusions

In summary, the current study observed age-related declines in facial expression and facial identity recognition abilities in a relatively large, healthy, age-diverse sample. Importantly, these declines were not fully explained by adjusting for the known age-related decline in general intelligence, even when this was thoroughly measured. Furthermore, the declines in expression and identity recognition were to some extent domain-specific, and not merely a function of broader face processing age-related difficulties.

Chapter 8: General Discussion

8.1 Thesis Overview

The research reported in this thesis investigated individual differences in a cluster of person perception abilities. Specifically, I have outlined the underlying structure of emotion expression recognition ability, and how this is associated with a range of broader cognitive and socio-affective traits. These included general intelligence, social cognition, facial identity recognition, autism-like characteristics, and alexithymia, as well as demographic variables of age and sex. A large body of research exists on the links of emotion expression recognition to broader life outcomes, including better social functioning and mental well-being. Equally, a difficulty in recognising emotional expressions has been shown to relate to the development and maintenance of psychiatric difficulties in a number of disorders. In this thesis, I examined expression recognition and identity recognition using an individual differences approach as a novel route to extend knowledge about these person perception abilities.

The introductory chapter first reviewed whether there is thought to be a general factor underlying the ability to recognise expression across emotional categories and sensory modalities. I also examined the current thinking on individual differences in expression and identity recognition, and the psychological and demographic factors that may contribute to these differences. There is little work addressing this question, and even less work using the powerful approach of structural equation modeling (SEM), and this chapter reviewed the research done thus far.

The empirical work presented in this thesis sought to address five main research aims. Firstly, I investigated the individual differences structure of emotion expression recognition ability. Secondly, I examined if this ability is related to broader abilities of face identity recognition and general intelligence. Thirdly, I assessed if general intelligence was related to the construct of alexithymia, after accounting for (facial and supramodal) emotion recognition ability. The fourth aim examined the long-held view of female superiority for emotion expression recognition. Finally, the fifth aim explored how face abilities change over adult age, and whether any change is independent of the known age decline in general intelligence.

The aim of this final chapter is to review in turn the findings of each paper, and consider how they relate to previous research in the field. Moreover, the wider implications are discussed, and future directions and conclusions are considered.

8.2 Chapter Summaries: Main Research Questions and Findings

8.2.1 Chapter 3: Emotion expression recognition ability: evidence for a supramodal factor and its links to social cognition

The first key research question addressed in this thesis was that of the underlying individual differences structure of emotion recognition ability. Much of the work concerning emotion expression recognition has focused only on facial expressions, despite expressive cues being conveyed through a variety of other modalities in everyday life. Theoretical perspectives posit that the ability to rapidly integrate cues across different modalities is adaptive for social interactions, and therefore it is plausible that individual differences across modalities reflect a higher-order general ability factor structure (Young, 2018).

To examine this, in Study 1 of Chapter 3, participants (N=284) completed a battery of previously piloted face, body, and voice expression tests, as well as measures of autism-like traits, social anxiety disorder and alexithymia. Using SEM, I tested a number of specified models, and found that the best fitting structure was a higher-order model with a superordinate general factor and three modality-specific factors (one for each of face, body and voice). In terms of the affective traits, there was a significant negative association between the supramodal emotion factor and alexithymia, but no relation to either social anxiety or autism-like traits. The lack of relation to autism traits is not wholly surprising given recent work that suggests co-occurring alexithymia is the reason for the reported emotional difficulties in many autistic individuals (Cook et al., 2013). Many studies have observed a link between social phobia and difficulty with recognising emotions, but this is not confirmed in Study 1 here. This may be in part due to the very brief three-item measures used, although it has reportedly good psychometric properties and high sensitivity for identifying socially anxious individuals (de Lima Osório, Crippa, & Loureiro, 2007). It should also be considered that the online sample tested here was healthier and more heterogeneous in nature than the clinical groups often tested. An association may exist but only at the more severe, clinical end of the social anxiety continuum.

In Study 2 (N=218), I extended the scope of the general emotion factor to include performance on a more ecologically valid test: the Reading the Mind in Films (RMFT) task. Given that this test comprises more realistic everyday social interaction scenarios, and also involves rapidly integrating expressive cues from characters' faces, bodies, and voices, this test is a pertinent measure to include as a test of the 'generality' of the general emotion factor. The RMFT performance significantly loaded onto the general factor, thus supporting the factor as accurately reflecting performance across a range of emotion tasks and modalities.

Additionally, participants completed a number of face processing tests, including unfamiliar face identity recognition. Results suggest the existence of a latent 'face perception' factor that is distinct from, yet highly related to, the general emotion recognition factor. These results are in line with recent theoretical perspectives of face processing, whereby facial identity and facial expression recognition share a common early perceptual stage. I modelled the relation between the general emotion and general face factors as a latent factor reflecting 'social cognition'. Future work examining a greater number and variety of tests will help to corroborate and extend the scope of this social cognition factor. It will be especially important to explore this ability factor by developing suitable batteries of non-face stimuli, given the relative reliance on facial stimuli in this study.

In Study 3 (N=249), a comprehensive measure of general intelligence was included in order to test its association to social cognition, and I also explored direct links to the general emotion and face latent factors. General intelligence was strongly and significantly predictive of the emotion factor, but not the face perception factor. This result is consistent with the perspective that face identity recognition is a specific cognitive ability, distinct from general cognitive ability. It also supports the proposed existence of a face perception '*f*' factor (Verhallen et al., 2017) that is separate from general cognitive ability.

Across three independent samples, then, I observed strong support for a general emotion recognition factor that extends beyond facial stimuli to include bodily expressions, auditory stimuli from the voice, and emotionally complex video stimuli. Given that this is also in line with, and in fact extends, the model structures reported in previous factor analytic and SEM studies, it seems a replicable and robust finding.

Some further findings of this section are important to note: I observed that the general emotion factor was highly associated with face perception, but the two abilities remained functionally distinct. Moreover, they showed differing relations to general intelligence. While emotion recognition was strongly and positively associated with general intelligence, face perception exhibited no association, in line with much research positing face perception as a distinct skill from general cognition. In contrast, accurate emotion recognition appears to reflect broader cognitive processes, and may be due to recruitment of problem-solving or working memory strategies.

8.2.2 Chapter 4: Recognition of facial expression and facial identity in part reflects a common ability, independent of general intelligence and visual short-term memory

Face perception literature has long debated over the degree of separation between identity recognition and expression recognition, and this paper sought to further address this question using an individual differences approach. Classic cognitive models have posited functionally distinct processing routes for facial cues that can rapidly change (i.e. emotional expression, eye gaze direction) and for cues that are relatively invariant (identity) (Bruce & Young, 1985; Haxby et al., 2000). Much of the support for this perspective arose from studies of neuro-damaged individuals who reported double dissociations between recognition of facial identity and facial expression (e.g. Young, Newcombe, de Haan, Small, & Hay, 1993).

However, more recent empirical work has contested the notion of complete functional segregation, and instead has posited some overlap between these processes. To address the question of degree of overlap, in Chapter 4, I analysed a large, healthy, age-diverse dataset (N=605) whom had been tested on well-used measures of unfamiliar facial identity recognition (Benton test of facial recognition) and facial expression recognition (Emotion Hexagon tests comprising morphed basic emotions on Ekman faces). I used a SEM approach, and observed a significant strong association between the two face abilities. In a second model, I examined this association whilst adjusting for the latent variable of general intelligence (Cattell culture-fair intelligence test) and a measure of visual short-term memory (a colour-wheel memory paradigm). Importantly, the expression-identity association, whilst reduced to moderate in magnitude, was still highly significant when accounting for performance on these two cognitive-perceptual measures. This suggests that the two face abilities share a common stage specific to face processing, and that cannot be fully accounted for by cognitive or perceptual abilities.

The findings are in line with Chapter 3 and previous work which indicated that individual differences in face perception are reflected at multiple levels, including general cognitive ability, broader face processing ability, and specific expression or identity recognition ability. Furthermore, results are in line with findings from Chapter 3 in that expression and identity recognition are related abilities but remain distinct in their own right. Finally, results appear to support the notion of the aforementioned *f*' factor, posited by Verhallen and colleagues (2017), that is distinct from other cognitive abilities.

In this study, a significant association between general intelligence and face identity recognition was observed, whereas no such association was observed between intelligence and the general face perception factor in Chapter 3. The discrepancy may be in part due to the choice of face perception task and the level of perceptual or cognitive demands each requires. In Chapter 4, the face test was the Benton test of facial recognition: a measure of unfamiliar face matching that can arguably be somewhat successfully completed by relying on low-level perceptual abilities. In contrast, the face tests used in Studies 2 and 3 of Chapter 3 comprised the Glasgow Face Matching Test and the Cambridge Face Memory Test, the latter of which (as its name suggests) has a large memory component. It thus requires higher-level face processing faculties to succeed, rather than lower-level feature matching abilities to problem-solve, as can be used to complete the Benton test.

8.2.3 Chapter 5: Alexithymia and General Intelligence: Independent association over and above emotion recognition ability

The sub-clinical construct of alexithymia has long been associated with emotional impairments. Individuals scoring highly on this construct report difficulties with describing and identifying their emotions, as well as a paucity of imagination and an externally-oriented style of thinking. The construct has been associated with a range of psychological and psychosomatic disorders, as well as with interoceptive ability, frontal lobe functioning, and verbal abilities, but it has not been well established whether the construct is related to general intelligence. This research question was robustly examined in Chapter 5. Across three independent, large, healthy samples, I observed a significant negative association between alexithymia and intelligence. Importantly, in two of the studies, this association remained significant when adjusting for emotion expression recognition ability. This suggests that the difficulties associated with alexithymia are broader than has typically been considered in the literature. It also suggests that the cognitive and emotional deficits observed in alexithymia are relatively independent of each other.

This finding has practical implications for the measure of alexithymia in clinical populations in which difficulties with cognition are already known, for example, depression (Marazziti et al., 2010). It is suggested in Chapter 5 that when assessing alexithymia and emotion recognition in these populations, a well-established measure of general intelligence should also be included to adjust for this potentially confounding variable.

One interesting outcome in this section was the finding that in two samples (Study 1a and Study 2), alexithymia was associated with both verbal and nonverbal components of the general intelligence measure, and the magnitudes of these associations did not significantly differ from each other. From the alexithymia-language hypothesis (Hobson et al., 2019) and from other studies implicating verbal ability in the alexithymia construct (Hsing et al., 2013), this result seems surprising. However, from the present results, it appears that alexithymia reflects broader, nonverbal aspects of cognition in addition to the verbal association that has been more consistently observed. This has theoretical implications for the alexithymia construct, and offers possible aetiological routes to its development. This is discussed in more depth below.

8.2.4 Chapter 6: Sex Differences in Emotion Recognition: Evidence for a Small Overall Female Superiority on Facial Disgust

This paper aimed to strongly test the widely held assumption of female superiority on emotion expression recognition. Early work in the field proposed the robust finding of a general female advantage across emotions, and more recent meta-analytic studies have supported the idea of females performing more accurately *overall* relative to males. However, other findings have reported a more mixed picture, especially with regards to individual emotion categories as opposed to a global score. Moreover, the vast majority of research had focused on facial expressions only, whereas a more extensive test of emotion recognition ability should incorporate expressions from other modalities. Further limitations of studies in the field consist of small underpowered samples, and prototypical expressions that often result in easy recognition and ceiling effects: these effects then obscure any potential sex differences that may exist.

In order to address these limitations, participants in Chapter 6: Study 1 were tested on the five basic emotions in static facial and dynamic bodily expressions in a large non-student adult sample (N=1022). The expressions had been carefully piloted prior to the study to eliminate items with floor or ceiling effects, thus making them suitable for individual differences work. Results indicated a significant female advantage for recognising facial disgust and facial sadness, and a significant male advantage for recognising bodily happiness.

With issues of reproducibility in mind, the results of the first study were tested in two further independent samples (Study 2: N=303, Study 3: N=634). In Study 2, participants completed the same battery of static faces and dynamic bodies as in Study 1, but with the additional batteries of dynamic faces and static bodies. Results confirmed the significant female advantage for static facial disgust that had been observed in Study 1. This finding was bolstered by a significant female advantage for dynamic facial disgust in the additional stimulus set, supporting the idea of a general ability across expressive modalities. A female advantage for static bodily happiness was also observed, but should be interpreted with caution given the male advantage noted in Study 1, and is most likely the result of random variability.

In Study 3, participants completed the static facial and dynamic bodily expression batteries as used in Study 1. Only the three significant findings that had been observed in Study 1 were tested, in order to increase power and validate only the robust findings. The significant female advantage for recognition of static facial disgust expressions was confirmed. The other two findings did not replicate.

Across three large independent samples then, I observed that females were consistently more accurate in recognition of facial disgust, and this ability extended beyond static expressions to incorporate dynamic stimuli as well. In contrast, a female advantage for facial sadness and a male advantage for bodily happiness did not replicate in either of the latter samples, and thus appear not to be robust findings. Overall, results of these three studies support the idea of a sex difference but it is important to emphasise that this difference is limited to facial disgust, and not any of the other basic emotions. Moreover, the effect, though significant, was small, and indicates that the two sexes are largely overlapping in their emotion recognition ability. This contrasts with previous work (e.g. Hall, 1978; McClure, 2000; Thompson & Voyer, 2014) and popular belief in the female advantage for emotion recognition that is often posited to be large and present across emotions, and instead suggests that for the most part, males and females are similarly accurate in recognising emotions.

8.2.5 Chapter 7: Face perception across the adult lifespan: Evidence for age-related changes independent of general intelligence

In the final empirical section, Chapter 7 addressed the question of how emotion recognition ability changes across the adult lifespan. Given the consistent finding of general intelligence declining with increasing age, and given also the strong positive associations between intelligence and emotion recognition that I have reported in Chapter 3 and Chapter 4, it is plausible that any age-related decline observed in emotion recognition may be reflective of the known decline in general intelligence.

To test this, a relatively large, healthy, age-diverse sample (N=595) completed widely used measures of facial emotional expression recognition, unfamiliar facial identity recognition, and general intelligence. Results indicated that both expression and identity recognition abilities declined significantly with age, independent of the decline in intelligence. Furthermore, the presence of both aspects of face perception meant that the specificity of the decline was explored. Of particular note, I observed that the decline in expression recognition was distinct from that of identity recognition, and vice versa. These findings are in line with the aforementioned perspective of face perception ability reflecting a hierarchical structure with multiple levels of individual differences (Chapter 3), and that at least part of the age effect appears to operate at the specific expression and identity levels. The age-related declines are considered relatively focal, and not *purely* due to declines in broader face processing or general cognitive ability.

8.3 Main Implications for Theory

8.3.1 The individual differences structure of emotion recognition ability

In Chapter 3, I outlined three independent studies that examined the functional architecture of emotion recognition. Results strongly suggested a general factor underlying recognition of emotions from across visual (face and body) and auditory (voice) modalities. As noted in the introductory and discussion sections of this chapter, this structure of emotion recognition may reflect the response to the demands of everyday social interactions. It is clear that individuals do not only communicate their emotional state through facial expressions, but also through other expressive cues, such as hand gestures, body posture, and vocal tone. It has been theorised in the literature that being able to rapidly attend to any or all of these cues may be the most adaptive strategy in accurately perceiving others' affective states. Indeed, a consistent finding is that of an advantage for bimodal (face and voice) stimuli: these are recognised more quickly and accurately than unimodal displays (Klasen, Chen, & Mathiek, 2012).

In the three studies in Chapter 3, the cues from sensory channels were not presented simultaneously (i.e. cross-modal). Instead, they were tested consecutively in separate batteries, and the common variance between each was extracted using a SEM approach. In this way, the individual differences structure of the general ability underlying performance across the three modalities was estimated free from measurement error.

Across the three studies in Chapter 3, the higher-order model exhibited the best fit to the data, and this is in line with recent work suggesting there exists multiple levels of individual differences within the architecture of emotion recognition ability, including modality-specific (i.e. face, body, or voice) and modality-general. It has been argued that there may be a good reason for the architecture to be arranged in this way. A general emotion factor that ranges across expressive modality may be particularly useful when the individual needs to integrate cues from multiple modalities at once, or when the expressive channels are changing rapidly. This is often the case for social interactions, where a person has to listen to what someone is saying, attend to their tone of voice, and perceive their facial expressions and/or body signals. All of these may change very quickly and subtly, and need to be perceived in parallel. A general factor in common cases such as this may be the most optimal solution for fast and accurate emotion recognition. Equally, modality-specific factors may be optimal when expertise for one expressive channel is especially required. Thus, the hierarchical structure outlined in Chapter 3 appears to reflect the full range of everyday social demands.

This study used structural modeling to look at face, body, and voice recognition in a single sample, and results support previous neuroimaging (Calder & Young, 2005) and structural equation modeling work (Schlegel et al., 2012) suggesting a general ability for perceiving emotional information from different sensory modalities. Importantly, results also support the previous SEM structures that posit a general recognition ability across emotions and across one (Suzuki et al., 2010) or two (Lewis et al., 2016) modalities, but extend this structure to three domains of face, body and voice. Given that this outcome was observed across three independent samples in Chapter 3, it appears a robust and replicable pattern.

8.3.2 Classic models of face perception

Classic cognitive models of face perception have posited functionally distinct theoretical and neural pathways for processing of facial expression and facial identity (Bruce & Young, 1986; Haxby et al., 2000) that diverge after a common stage in face perception. However, more recent work has indicated that variant (expression, gaze) and invariant (identity, sex) facial signals may interact to a greater degree than was previously assumed. Becker and colleagues (2007) observed an interaction between sexually dimorphic characteristics and expressions of anger or happiness, such that males were more likely to be labelled as displaying an angry expression and females as displaying a happy expression. Responses were also shown to be quicker when these two types of facial signal were congruent (i.e. anger/male and happiness/female) as opposed to incongruent (anger/female and happiness/male). In a similar vein, Sacco and Hugenberg (2009) found an interaction between the invariant signal of face shape and the variant signal of fear and anger expressions. They observed an interaction effect such that a more youthful and submissive-looking face (rounder face, higher brow, larger eyes) was perceived to be showing a fearful expression, whilst a more mature, dominant-looking face (narrower face, lower brow, smaller eyes) was perceived as expressing anger.

Given some evidence for an overlap between expression and identity, then, an interesting question concerns the extent of this common stage of shared processing. In Chapter 4, a strong positive association between facial expression and facial identity was observed (.52), indicating a substantial degree of overlap between the two face perception abilities. This suggests that the two measures included here are tapping abilities at the early perceptual stage of face processing, rather than a later cognitive stage. In Studies 2 and 3 of Chapter 3, the relationship between facial expression recognition and the two tests of unfamiliar face recognition (GFMT, CFMT) was weaker (between .19 and .30), indicating a lower degree of overlap. This makes sense, given that the CFMT in particular places fewer demands on earlier perceptual processes and more emphasis on later cognitive processes. From these studies, it is clear that the degree of overlap between expression and identity recognition is more nuanced than traditional models posited, and depends largely on the specific demands of the tasks used.

An important observation from this section is that the expression-identity association was reduced in magnitude but remained significant when adjusting for measures of general intelligence and visual short-term memory. This implies that at least some of the shared variance arises from the level of specific face processing, and that it is not fully explained by a commonality at the level of broader cognitive processing. These findings are consistent with previous work (Lewis et al., 2016) and with results from Chapter 3 of the current thesis that face processing involves contributions from individual differences at multiple levels within the ability, including general cognition, general face processing, and emotion- or identity-specific processing.

Results can be interpreted in line with both Bruce and Young's and Haxby et al.'s models, given that some of the overlap appeared to be at the cognitive level, and some appeared to be at a face-specific level. Determining the structure of multiple levels involved in face processing helps us to understand the discrepancies in the literature regarding the extent of the overlap between facial expression and identity recognition.

8.3.3 The *f* factor in face perception

Across many broad batteries of general intelligence, a commonly reported finding is that of positive correlations being observed between individual sub-tests, otherwise known as the 'positive manifold', and the term *g* was coined to describe the general factor that emerges through factor analysis of the batteries (Spearman, 1927). In the same vein, Verhallen and colleagues (2017) observed significant correlations across a variety of face perception tests, and thus termed the underlying common factor *f* that accounted for the positive manifold performance across these tasks. Other studies have observed similar results for significant inter-relations between diverse tests of face processing (Bowles et al., 2009; Wilhelm et al., 2010; McCaffery, Robertson, Young, & Burton, 2018). In both Chapter 3 and Chapter 4, results were in line with these studies and appeared to support the existence of this general face perception factor *f* that was related to but functionally distinct from other cognitive and socio-emotional abilities.

This proposal of *f* was particularly supported by results of Chapter 3, given that the same three face perception tests (Glasgow Face Matching Test, Cambridge Face Memory Test, Mooney test) were included as in Verhallen et al.'s *f* factor paper. Moreover, this chapter contributed to extending the scope of *f* by showing that it relates to, but remains distinct from, other higher-level cognitive abilities including facial expression recognition. Its scope was also extended by observing support for the *f* factor in a broader, more diverse and more generalisable sample than the undergraduates tested by Verhallen et al., and across two independent studies.

Verhallen and colleagues reported no significant association between performance on the Cambridge Face Memory Test (CFMT) and participants' scores on the UK secondary school qualification: GCSE (their proxy for intelligence). In contrast, I observed a significant positive correlation between the Benton Test of Facial Recognition and the Cattell Culture-Fair intelligence test in Chapter 4. This overlap may be as a result of similar cognitive strategies that are used in the multiple choice general intelligence test and can also be employed to 'solve' the Benton test. Studies have shown that typical scores on this test can be achieved even by prosopagnosic individuals, by means of matching the facial features in a non-holistic style (Bowles et al., 2009). In this way, participants can achieve 'normal' scores on then test even in the absence of 'normal' face recognition ability. In contrast, the CFMT does not show the target and test faces simultaneously and so the participant is required to encode the target identity in memory in order to succeed on the test, and this memory component for encoding faces may show less commonality with the intelligence test in Chapter 3.

Alternatively, the lack of correlation between GCSE scores and CFMT performance may have been as a result of the undergraduate sample drawn from the same university. Verhallen himself notes that this may have restricted the possibility of observing a correlation given that many of the students may have achieved similar GCSE grades, thus limiting the range extensively. One strength of the methods used is Chapter 4 is the inclusion of a battery of general intelligence that tested participants at the same time point as performance on the face perception tasks. This is in contrast to the inclusion of GSCE scores which may have been sat a few years prior to the face battery tests, and which required participants to remember and report their scores accurately and truthfully.

The strong correlation between facial expression and facial identity recognition observed in Chapter 4 lends support to the notion of a general face factor. However, it is also possible that the association between these two particular measures may have been somewhat bolstered by the overlap in perceptual demands of the two face tests. Both expression labelling and unfamiliar face matching tasks demand contribution from early perceptual processes, and as highlighted previously, it is likely that a weaker correlation would have been observed in this study had a test like the CFMT been used, in which later cognitive processes are more pertinent to task success. This is supported by the findings from Chapter 3 in which a weaker association between the CFMT and facial expression recognition ability was observed.

Nonetheless, the association between the expression and identity measures in Chapter 4 remained significant and moderate in magnitude when adjusting for cognitive and perceptual ability, thus giving support to the face-specific *f* factor that is distinct from wider cognitive functioning. The moderate association between the abilities, then, appears to better capture the face-specific variance of the general *f* factor.

8.3.4 Theories of alexithymia

Verbal difficulties have frequently been reported in alexithymic individuals, and a recent proposal has extended this perspective in greater depth. In their alexithymia-language hypothesis, Hobson and colleagues (2019) suggested that early language difficulties may be one plausible route for development of alexithymia, given that both nativist and constructionist theories of emotion posit language abilities as critical for emotional development. According to this model, then, any dysfunction of these language abilities could have a disruptive effect on subsequent emotional abilities (Hobson et al., 2019).

However, Hobson and colleagues acknowledge that verbal impairment is not the only possible route to alexithymia. Other potential mechanisms include interoceptive difficulties (Brewer, Cook, & Bird, 2016), cortical damage, as posited by the neurobiological accounts (Larsen, Brand, Bermond, & Hijman, 2003) or difficulties focusing on or evaluating an emotional response, as posited by the attention-appraisal model (Preece et al., 2017). In the samples tested in Chapter 5, alexithymia was negatively associated with both verbal and nonverbal components of the general intelligence test, suggesting that more general cognitive deficits appear to exist. Cortical lesions to frontal or right cerebral regions may contribute to these broader impairments, given the cognitive processes that are associated with frontal lobe functioning, such as working memory, decision-making, and task-switching. According to neurobiological perspectives (e.g. Larsen et al., 2003), deficits to these areas would result in neither experience of emotion (affective component) nor subsequent conscious reasoning of emotion (cognitive component). Rather, a relative blunting of emotionality is thought to occur.

In contrast, alexithymia arising through interoceptive impairment is likely reflecting a type of alexithymia in which emotions *are* experienced but unable to be explicitly verbalised. Because of this, these individuals may report more psychosomatic problems, arguably due to a misinterpretation of the somatic symptoms that come from emotional arousal (Larsen et al., 2003). This is consistent with interoceptive alexithymia, in which physiological signals are unable to be interpreted appropriately and often misconstrued for other internal cues. Finally, Preece et al.'s (2017) attention-appraisal model posits that an individual experiences an emotional response but because of their under-developed emotion schemas, they either do not focus on the most pertinent aspect of the response or they misinterpret the emotional information, resulting in emotional difficulties and maladaptive emotion regulation.

These possible routes for development of alexithymia (frontal damage, interoceptive failure, or underdeveloped emotion processing) may also account for the heterogeneity observed in the many psychosomatic and psychological disorders associated with the construct. If an individual has developed alexithymia through the frontal lobe/right hemisphere pathway, then they have an absence of both the emotional and cognitive aspects of affect, and as such, will not suffer problems with emotionalising or psychosomatic distress. In contrast, individuals with impaired interoception continue to experience emotional arousal but with no conscious cognition, and therefore, the physiological signals can be misinterpreted. As a result, individuals have problems regulating their internal bodily signals and emotions, leading to a variety of psychosomatic or emotion dysregulation conditions, such as anorexia nervosa (Bourke, Taylor, Parker, & Bagby, 1992; Brewer et al., 2015), chronic pain (Lumley, Asselin, & Norman, 1997), alcohol abuse (Lyvers et al., 2012), and medically unexplained symptoms (Deary, Scott, & Wilson, 1997), amongst many others. In a similar vein, if an individual's alexithymia is as a result of poorly developed emotion schemas, then they experience an emotional response but do not attend to or evaluate it appropriately, resulting in poor regulation of uncomfortable feelings and negative emotional reactivity (Preece et al., 2017).

In terms of the alexithymia-language hypothesis, Hobson et al. (2019) noted that it is difficult to distinguish between alexithymia caused by damage to language processing areas such as the inferior frontal gyrus, or damage to interoception-related areas such as the anterior insula, given how close in proximity these areas are to each other and their location being close to or within the frontal lobes themselves. It may be that the pathways outlined in the different manifestations of alexithymia are not mutually exclusive, and further behavioural and structural imaging research could help to disentangle their respective aetiologies.

As mentioned, one key finding from the studies in Chapter 5 is that alexithymia is not solely related to verbal difficulties but shows associations with broader cognition too. In Preece et al.'s (2017) attention-appraisal model, the authors outline a route to alexithymia whereby an individual's emotion schemas are underdeveloped and not well organised or integrated. As first posited by Lane and Schwartz (1987), emotion schemas are cognitive structures that guide the processing of emotions. If these have not developed normally, it results in the individual reaching only the lower-level basic stages of emotional awareness, with less complex and nuanced processing of emotions. The underdevelopment of these key cognitive structures may imply some broader cognitive development difficulties in the individual as well.

Izard defined emotion schemas as "emotion feelings interacting with cognition in motivating the decision making and actions of everyday life" (Izard, 2011, p371). He argued that emotion schemas depend on sufficient cognitive development in order to implement the higher-order mental processes involved in emotional experience. Given that Izard suggested that the basic emotions require minimal cognitive processing relative to more complex secondary emotions, and that alexithymia was associated with inferior recognition of the basic emotions in the studies in Chapter 5, it would be interesting in future studies to examine whether people higher in alexithymia would show even worse recognition of these more complex and blended emotions.

In sum, a possible route to alexithymia is through inadequate development of language and/or emotion schemas that lead to inappropriate emotion appraisal. These schemas are dependent on adequate cognitive development, especially higher-order processes including reasoning, planning, and evaluation of complex information. The manifestation of emotion recognition difficulties as seen in the three samples in Chapter 5 may reflect insufficient cognitive development, and may be one plausible explanation for the observed difficulties in general intelligence in people high in alexithymia observed here.

8.3.5 Sex differences in emotion recognition ability

It has been widely reported in emotion literature to date that females consistently show a general advantage over males for accurate recognition of emotional stimuli. Many individual empirical studies and four meta-analyses (Hall, 1978, 1984; McClure, 2000, Thompson & Voyer, 2014) have supported this finding, and report small-to-medium effect sizes. Hall's first review indicated a Cohen's *d* of 0.40, and her second indicated a range between 0.18 and 0.52. McClure's review (which it should be noted comprised children and adolescent studies) reported a *d* of 0.13, and Thompson and Voyer reported an overall *d* of 0.19. However, closer inspection of the available studies reveals a more nuanced picture with a greater mixture of outcomes. Whilst many studies report more accurate performance in females, a number of studies have observed no significant sex differences or report a male advantage.

The findings from Chapter 6 sought to offer some clarity with regards to this mixed literature. Across three relatively large, independent samples, a small but significant advantage for females on recognising facial disgust was observed (Cohen's *d* ranged from 0.18 to 0.47). Interestingly, this finding extended from static facial stimuli to a novel set of dynamic faces. Despite some sex differences observed in other emotional categories (for example, male advantage for bodily happiness, and female advantage for facial sadness), none of these replicated in the other two studies, and thus are most likely the result of random sampling. These results emphasise the importance of replication studies and validating results in well-powered samples.

The fact that the female advantage for facial disgust recognition was observed across the three studies in Chapter 6, as well as the advantage extending to include dynamic morphed facial stimuli, suggests that this is a robust pattern. However, it should also be highlighted that the effect sizes were small-to-moderate, in line with the magnitudes reported in the previous meta-analyses. Given the results from Chapter 6, then, it can be tentatively concluded that there does appear to be a robust sex difference, but importantly, it is small in magnitude and limited to the domain of facial disgust. In terms of global emotion recognition scores, there appears to be no sex difference, and overall, the distributions of the two sexes are largely overlapping.

In the discussion of Chapter 6, a speculative account was proposed for the robust female advantage for facial disgust recognition, and which focused on the distinct evolutionary pressures faced by males and females. Females have been shown to score higher on measures of disgust sensitivity, and this is considered to serve as a guard against disease. They have a higher risk of contracting sexually transmitted diseases and of passing infection on to any offspring. Their immunity changes over the first half of the luteal phase of the menstrual cycle when conception is most likely, and also during the first trimester of pregnancy when their maternal immune system is suppressed to prevent attack on the growing foetus (Fessler, Eng, & Navarrete, 2005). Food-borne illnesses are especially dangerous to the mother and unborn child during this initial pregnancy phase, which may in part account for the increased levels of nausea and vomiting, and may elevate reactions to subtle oral or vocal cues of food-related elicitors of disgust. Experiencing higher levels of disgust, then, may be a motivating factor in avoiding potential contaminants. From an evolutionary adaptationist framework, it is argued that this may drive the more sensitive and rapid recognition of disgust cues often reported in females.

Whilst females' reporting of higher disgust sensitivity may arise from hormonal influences, it has been argued that males' consistently lower scores on this measure is as a result of highlighting their robust immune systems (Fleischman, 2014). The selection pressure faced by males trying to attract and secure a reproductive mate may mean that it is not adaptive for them to outwardly react to signals of disgust or contamination. In fact, they may be more successful in acquiring a mate by expressing a willingness to approach potential toxins, as with other risky behaviour (Fessler, Pillsworth, & Flamson, 2004). By not recognising or reacting to subtle cues of disgust, males signal to potential female mates that their immune system is (and thus their genes) are strong and suitable for successful mating.

It is interesting that the female advantage for recognising disgust was limited to the facial domain in the studies of Chapter 6, albeit in both static and dynamic formats. This may be because disgust is primarily an emotion to protect against oral ingestion of toxins. Given that reactions to contaminants are most often expressed via the nose and mouth, it seems most adaptive to focus on these facial cues when scrutinising for signs of contamination. As the traditional primary caregiver in most instances, it is often the female's responsibility to carefully monitor their offspring for these signs and may result in their superior facial disgust recognition ability. Traditionally, males have a lower parental investment in raising offspring, and are not often the primary caregiver, and so sensitivity to their facial signals of disgust may not be salient or adaptive for males.

Another reason why the sex difference was found only for facial stimuli is that disgust is relatively difficult to express via the body alone, evidenced by the low performance scores of both males and females on the dynamic bodily disgust point-light displays in Study 1. In fact, both sexes scored the lowest on this domain. In addition, items expressing disgust through a static body were not included in the stimulus set that was used in Study 2, due to it being highly difficult to accurately portray disgust in this format (de Gelder & Van den Stock, 2011). In terms of disgust recognition, then, it appears most adaptive to focus on facial cues, although distinct mating roles may drive this ability differently in males and females. In sum, it is suggested, with a certain degree of caution, that unique selection pressures are a potential explanation for the observed small female superiority on facial disgust.

The sex difference in disgust recognition observed in Chapter 6 is in line with Darwinian and nativist theories of basic emotions outlined in the introductory chapter. It seems plausible that sex-specific selective pressures can drive experience and expression of the basic emotions differently in males and females, and this may vary according to each of the emotions and their adaptive functions. For disgust specifically, differences in child-rearing roles may have driven higher sensitivity and more accurate recognition ability in females, as this is advantageous for optimal offspring survival, whereas males are afforded no evolutionary advantage for recognition of facial disgust. Rather, males may face a selective pressure to *not react* to signals of contamination or infection, as a means of highlighting their reproductive fitness to females. If this is the case, it is perhaps not surprising that there exists a sex difference in favour of females on facial disgust. For the other four basic emotions, however, mating roles may not be as salient in shaping the respective selection pressures and may be why no robust sex differences emerged for these categories.

8.4 Critical Evaluation: Limitations and Future Directions

Individual limitations of each study have been discussed in the corresponding chapters. In this section, some general limitations that are relevant across the range of studies are considered.

8.4.1 Cross-modal versus supramodal emotion recognition

One clear limitation of the current work is that participants were not tested on emotional stimuli from different categories simultaneously. That is, *cross-modal* emotion recognition was not examined here. Rather, participants in Chapter 3 were shown emotional stimuli from the face, body, and voice separately and in succession, and the commonality between their performances was statistically extracted using SEM. This then resulted in a general emotion recognition factor that, along with my collaborators, we have termed ‘supramodal’. This factor reflects an individual’s ability to recognise emotions from a variety of visual and auditory sources, and offers interesting insight into the question of whether our ability to recognise facial emotions is associated with our ability to recognise cues from other sensory channels.

Of course, this approach does not assess an individual's ability to *integrate* across these channels, and future work using cross-modal stimuli will be invaluable towards establishing the nature of sensory integration of emotional information. Currently, there exists few available stimulus sets to robustly examine this which are suitable for individual differences work, i.e., stimuli that show no floor or ceiling effects, and are sensitive enough to detect a wide range of individual variation within a healthy sample. Development and validation of such sets is an ongoing challenge for the field, and a key aim for future research. Given the interesting results reported in Chapter 3, such stimuli development would be valuable for exploring emotion recognition ability from a cross-modal perspective.

Another reason why future testing of cross-modal emotional stimuli is particularly relevant is the higher ecological validity they can offer. It has been widely argued that in everyday life, we do not often encounter emotional stimuli in isolation. When interacting with other people, we can often see their facial expressions and hear their tone of voice simultaneously, as well as perceiving their hand gestures and body posture. Being able to quickly and accurately attend to all of these various sensory cues in parallel is fundamental for a nuanced understanding of the emotional state of the other person and for responding most appropriately. From an evolutionary perspective, congruent vocal, facial and bodily expressions may have facilitated the rapid recognition of a signal of threat in the environment or the state of mind of a conspecific.

For the subsets of participants who show relative difficulties with emotion recognition, it may be useful to have access to other emotional cues as a compensatory strategy for decoding expressions. For example, if a person who is higher in alexithymia is unsure of a facial expression being made, the accompanying body language or vocal tone may help them to work out the most likely emotion being expressed. Alternatively, bi- or trimodal stimuli may offer the additional insight of identifying which sensory channels certain subsets of participants more heavily rely on for recognition. For example, in their Reading the Mind in Films task (RMFT), Golan and colleagues (2006) demonstrated that autistic participants were less likely to integrate the verbal information with the available prosodic and facial cues, thus resulting in a higher proportion of labelling errors. The authors recommend the development and use of ecological, multisensory tasks to assess socioemotional understanding that may reveal more subtle and nuanced difficulties in some clinical populations. To date, the widespread use of static, posed facial expressions of emotion may be easy enough for autistic or alexithymic participants to pass, and therefore their emotional difficulties and recognition strategies are somewhat masked by the choice of stimuli.

Older adults are also thought to be able to somewhat compensate for their declining emotion recognition by taking into account emotional signals from multiple sources. In a study of lexical, facial, and vocal expressions, older adults benefitted more from bimodal or trimodal stimulus presentation than did their younger counterparts. There was no group difference between the younger or older participants on the bi- or trimodal conditions, but the older adults showed significantly worse accuracy on unimodal expressions (Wieck & Kunzmann, 2017). Consistent with this notion, some studies have reported equal or even superior performance of older adults when the emotional stimuli consisted of videos rather than static images (e.g. Murphy & Isaacowitz, 2010). Isaacowitz and Stanley (2011) went on to argue that most emotion stimuli used to date have been impoverished and lacking in context. He argues that whilst facial emotion recognition does appear to decline in older adults, the reality of this decline may not be as severe as results would suggest. In everyday life, older adults may be able to effectively compensate for this decline by considering situational cues or signals from other sensory channels. Appreciating this compensation may lead researchers to a more nuanced understanding of the mechanisms underpinning age-related decline in emotion recognition (Isaacowitz & Stanley, 2011).

Whilst this ageing and emotion recognition literature remains somewhat mixed (Ruffman, 2011), it is plausible that older adults' emotion recognition ability may be enhanced with the addition of congruent information, and therefore cross-modal stimuli may minimise the age effect that is widely reported in the literature. It is possible, then, that the age decline observed in Chapter 4 would have been weaker had the emotional stimuli comprised simultaneous bodily or vocal information alongside facial expressions. Future studies with similarly age-diverse samples and cross-modal stimuli will be able to offer a unique insight into this debate.

8.4.2 Testing in general populations

Another limitation that is relevant across the studies presented in this thesis is the testing of clinically-relevant traits in healthy populations. Whilst the findings are informative for exploring the associations between emotion expression recognition and socio-affective disorders, it does not necessarily follow that these associations translate to a clinical population.

For example, in Chapter 3, I tested autism-like characteristics using a 28-item self-reported questionnaire. Whilst this instrument provides a range of individual variation and is useful in establishing correlations with emotional tasks, it is obviously not as thorough as a clinical autism assessment and interview in identifying the severity of autistic traits. Furthermore, autism-like traits were tested in an online sample of healthy American adults. Online samples have been demonstrated to be more diverse and more representative of the general population at large than traditional college-based samples, but of course are not representative of clinically-diagnosed autistic populations. Confounds of higher general intelligence and verbal ability in the healthy samples in Chapter 3 may have constrained or altered the associations reported. Future work examining the structure of emotion recognition ability in autistic individuals may corroborate the superordinate emotion factor reported here, or alternatively may reveal a different underlying structure in this population.

Also in Chapter 3, social anxiety was measured and no association with emotion recognition was observed. However, for purposes of overall survey brevity, the measure used was very short and may not have reliably captured the variance of the disorder. Moreover, as with autism-like traits above, the samples tested were healthy, and therefore it cannot be ruled out that an association with emotion recognition would have emerged if a clinically diagnosed sample of individuals with social anxiety disorder had been tested.

Finally, in both Chapters 3 and 5, the sub-clinical construct of alexithymia was assessed using the self-report Toronto Alexithymia Scale. Whilst this scale is very widely used in the field, it has faced criticisms regarding the self-report nature. Some researchers argue that being able to provide an accurate and reliable gauge of emotional capabilities requires a high degree of self-awareness that alexithymics typically do not have (Hobson et al., 2019). This lack of meta-awareness into one's own emotional difficulties may confound the measurement of alexithymia, particularly in high-scoring individuals who may not be aware of their own impairments. Furthermore, verbal difficulties in alexithymic individuals may interfere with their self-assessment. By rating agreement to items like 'It is difficult for me to find the right words for my feelings", the researcher may be unsure as to whether the problems lie in *identifying* the emotional feeling, or in *verbalising* their emotional experience. However, at least in the healthy general populations tested, alexithymia scores correlate well with emotional task performance, suggesting that the scale does have predictive value and that most individuals do have reliable emotional insight. As with the two clinical traits discussed above, it would useful for the understanding of the alexithymia construct if the associations reported here were corroborated in individuals who score above the clinically-significant threshold for alexithymia.

8.4.3 Social Cognition variable – reliance on facial stimuli

The overriding majority of the current set of studies has relied on facial stimuli for measurement of social cognition, and this reliance may be limiting. In Chapter 3, a latent social cognition factor was proposed, but the subtests included in the overall model comprised mainly facial stimuli. It would have been interesting to see if the latent face perception factor (derived from face identity recognition tasks) would have extended to identity recognition from other modalities, for example, voice. However, body and voice identity recognition tasks that are suitable for individual differences research are not widely available. In pilot studies of voice identity tests carried out during this doctoral work, the psychometric properties observed were not satisfactory for the individual differences approach that was used with the other stimulus sets. That is, in two voice identity recognition tasks tested, participants showed chance level responses and little individual variation. The lack of stability of these two tests rendered them impossible to structurally and stably model, and therefore could not be used to answer the theoretically interesting question of whether the face perception factor extends to other sensory modalities, (e.g. voice, gait), and better reflects ‘person perception’, rather than faces *per se*. As highlighted, a pertinent task for future research in this area is the development and validation of suitable stimulus sets that will help to broaden the scope of the latent ability factors observed in this thesis.

Regarding emotion expression recognition, the development of more ecologically valid stimuli will be of great use to the field. Stimuli that vary in race, age, and modality, as well as showing a wider variety of complex emotional expressions and social interactions, will be necessary for corroborating and extending the findings reported in this thesis. As outlined in previous chapters, there has arguably been an over-reliance on greyscale posed expressions. Stimuli that present emotional expressions across modalities simultaneously may be especially helpful in detecting subtler individual differences in certain samples, for example, older individuals or individuals with autism. Within the general adult population, expressions that are not prototypical and posed, i.e. subtle, spontaneous micro-expressions, are required to avoid the widespread ceiling effects in the field, and enable the researcher to observe the more nuanced differences in expression recognition ability.

8.5 Conclusions

In sum, this thesis has advanced knowledge of person perception abilities in several important ways. Firstly, I have outlined the individual differences structure of emotion recognition ability across face, body and voice domains, and suggest that this structure is a higher-order model comprising a general supramodal factor and domain-specific factors. Secondly, I have extended our understanding of this structure by exploring its links to important cognitive and affective traits; specifically, I observed moderate links with general intelligence and alexithymia.

Thirdly, I also modelled another core person perception ability, that of face identity recognition, and observed that this ability was related to, but functionally distinct from, that of emotion expression recognition. Fourth, I used an individual differences approach as a novel and more robust way to examine how demographic variables of sex and age are associated with person perception abilities. These sections highlight the importance of ensuring that emotion and identity stimuli are suitable for detecting subtle individual variation, given how varied the healthy adult population appears to be on these abilities.

I have outlined the architecture of person perception abilities, and have also placed these abilities in a wider cognitive and psychosocial context. To gain a fuller and more nuanced understanding of emotion and identity recognition, it is important to understand how these abilities vary in a neurotypical population, as well as how they relate to broader everyday skills. The findings presented here emphasise how individual variation has been somewhat overlooked in earlier studies, and how it is crucial to consider when examining these abilities in future work.

References

Addington, J., & Addington, D. (1998). Facial affect recognition and information processing in schizophrenia and bipolar disorder. *Schizophrenia Research*, *32*(3), 171–181. https://doi.org/10.1016/S0920-9964(98)00042-5

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, *13*(2), 232-240.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669-672.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., ... & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, *37*(10), 1111-1117.

Aglieri, V., Watson, R., Pernet, C., Latinus, M., Garrido, L., & Belin, P. (2017). The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. *Behavior Research Methods*, *49*(1), 97-110.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory, (1973)*, 267–281.

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: An investigation of gender differences. *PloS One*, *6*(6), e20989.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion Perception from Dynamic and Static Body Expressions in Point-Light and Full-Light Displays. *Perception*, *33*(6), 717–746. https://doi.org/10.1068/p5096

Attwood, A. S., Easey, K. E., Dalili, M. N., Skinner, A. L., Woods, A., Crick, L., … Munafò, M. R. (2017). State anxiety and emotional face recognition in healthy volunteers.

Open Science, *4*(5), 160855. https://doi.org/10.1098/rsos.160855

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225-1229.

Bagby, M., Taylor, G. J., & Ryan, D. (1986). Toronto Alexithymia Scale: Relationship with personality and psychopathology measures. *Psychotherapy and Psychosomatics*, *45*(4), 207-215.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Baksh, R. A., Abrahams, S., Auyeung, B., & MacPherson, S. E. (2018). The Edinburgh Social Cognition Test (ESCoT): Examining the effects of age on a new measure of theory of mind and social norm understanding. *PloS one*, *13*(4), e0195818.

Bal, E., Harden, E., Lamb, D., Van Hecke, A. V., Denver, J. W., & Porges, S. W. (2010). Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of Autism and Developmental Disorders*, *40*(3), 358-370.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5-17.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *Journal of Neuroscience*, *30*(7), 2414-2417.

Beaujean, A. A., & Sheng, Y. (2010). Examining the Flynn effect in the General Social Survey Vocabulary test using item response theory. *Personality and Individual Differences*, *48*(3), 294-298.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*(2), 179-190.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. https://doi.org/10.3758/BRM.40.2.531

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, *42*(5), 825-829.

Benton, A. L., Eslinger, P. J., & Damasio, A. R. (1981). Normative observations on neuropsychological test performances in old age. *Journal of Clinical and Experimental Neuropsychology*, *3*(1), 33-42.

Benton, A. L., & Van Allen, M. W. (1968). Impairment in facial recognition in patients with cerebral disease. *Cortex*, *4*(4), 344-358.

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience*, *35*, 1-23.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, *3*(7), e285.

Bird, G., & Viding, E. (2014). The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, *47*, 520-532.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain*, *133*(5), 1515-1525.

Blair, R. J. R., & Curran, H. V. (1999). Selective impairment in the recognition of anger induced by diazepam. *Psychopharmacology*, *147*(3), 335-338.

Borod, J. C., Pick, L. H., Hall, S., Sliwinski, M., Madigan, N., Obler, L. K., … Tabert, M. (2000). Relationships among Facial, Prosodic, and Lexical Channels of Emotional Perceptual Processing. *Cognition & Emotion*, *14*(2), 193–211. https://doi.org/10.1080/026999300378932

Bourke, M. P., Taylor, G. J., Parker, J. D., & Bagby, R. M. (1992). Alexithymia in women with anorexia nervosa: A preliminary investigation. *The British Journal of Psychiatry*, *161*(2), 240-243.

Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Memory & Cognition*, *34*(4), 854-864.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., … Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455. https://doi.org/10.1080/02643290903343149

Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: a comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology*, *91*(4), 780-795.

Brewer, R., Cook, R., & Bird, G. (2016). Alexithymia: a general deficit of interoception. *Royal Society Open Science*, *3*(10), 150664.

Brewer, R., Cook, R., Cardi, V., Treasure, J., & Bird, G. (2015). Emotion recognition deficits in eating disorders are explained by co-occurring alexithymia. *Royal Society Open Science*, *2*(1), 140382.

Brownlow, S., Dixon, A. R., Egbert, C. A., & Radcliffe, R. D. (1997). Perception of movement and dancer characteristics from point-light displays of dance. *The Psychological Record*, *47*(3), 411-422.

Bruce, V., & Young, A. W. (2012). *Face perception*. London, Psychology Press.

Bruce, V. (1994). Stability from Variation: The Case of Face Recognition the M.D. Vernon Memorial Lecture. *The Quarterly Journal of Experimental Psychology Section A*, *47*(1), 5–28. https://doi.org/10.1080/14640749408401141

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

Buck, R., Miller, R. E., & Caul, W. F. (1974). Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology*, *30*(4), 587-596.

Buhrmester, M. D., Kwang. T., and Gosling, S. D. (2011). Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science 6,* 3–5.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*(8), 1467–1485. https://doi.org/10.1080/17470218.2013.800125

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42*(1), 286–291. https://doi.org/10.3758/BRM.42.1.286

Button, K., Lewis, G., Penton-Voak, I., & Munafò, M. (2013). Social anxiety is associated with general but not specific biases in emotion recognition. *Psychiatry Research*, *210*(1), 199-207. https://doi.org/10.1016/j.psychres.2013.06.005

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651. https://doi.org/10.1038/nrn1724

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*(9), 1179–1208. https://doi.org/10.1016/S0042-6989(01)00002-5

Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., & Young, A. W. (2003). Facial expression recognition across the adult life span. *Neuropsychologia*, *41*(2), 195-202.

Calder, A. J., Keane, J., Young, A. W., Lawrence, A. D., Mason, S., & Barker, R. A. (2010). The relation between anger and different forms of disgust: Implications for emotion recognition impairments in Huntington's disease. *Neuropsychologia*, *48*(9), 2719–2729. https://doi.org/10.1016/j.neuropsychologia.2010.05.019

Calder, A. J., Young, A.W., Rowland, D., & Perrett, D.I. (1997). Computer-enhanced emotion in facial expressions. *Proceedings of the Royal Society of London, B. 264*(1383), 919-925.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology*, *10*(11), 649-657.

Cangur, S., & Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. *Journal of Modern Applied Statistical Methods*, *14*(1), 152-167.

Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal Decoding Skills and Relationship Well-Being in Adults. *Journal of Nonverbal Behavior*, *23*(1), 91-100.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156-2160.

Cattell, R. B. (1973). *Culture-fair intelligence test, Scale 1, Handbook*. Institute of Personality and Ability Testing.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112-130.

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2019). Sex differences in emotion recognition: Evidence for a small overall female superiority on facial disgust. *Emotion*, *19*(3), 455-464.

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2020). Emotion recognition ability: evidence for a supramodal factor and its links to social cognition. *Cognition, 197*. 104166.

Connolly, H. L., Young, A. W., & Lewis, G. J. (2019). Recognition of facial expression and identity in part reflects a common ability, independent of general intelligence and visual short-term memory. *Cognition and Emotion*, *33*(6), 1119-1128.

Connor, K. M., Kobak, K. A., Churchill, L. E., Katzelnick, D., & Davidson, J. R. T. (2001). Mini-SPIN: A brief screening assessment for generalized social anxiety disorder. *Depression and Anxiety*, *14*(2), 137–140. https://doi.org/10.1002/da.1055.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science, 24*(5), 723-732.

Correro II, A. N., Paitel, E. R., Byers, S. J., & Nielson, K. A. (2019). The role of alexithymia in memory and executive functioning across the lifespan. *Cognition and Emotion*, 1-16.

Crivelli, C., & Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends in Cognitive Sciences, 22*, 388-398.

Curtis, V., Aunger, R., & Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease. *Proceedings of the Royal Society of London, B. 271*, 131-133.

Darwin, C. (1872/1965). *The expressions of the emotions in man and animals.* Oxford University Press, New York.

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. PloS one, 6(12), e28800.

de Gelder, B. & Van den Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology, 2*(181), https://doi.org/10.3389/fpsyg.2011.00181

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467. https://doi.org/10.1016/j.tics.2003.08.014.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*(3), 289–311. https://doi.org/10.1080/026999300378824

de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neuroscience Letters*, *260*(2), 133–136. https://doi.org/10.1016/S0304-3940(98)00963-X

De Gucht, V., & Heiser, W. (2003). Alexithymia and somatisation: a quantitative review of the literature. *Journal of Psychosomatic Research*, *54*(5), 425-434.

de Lima Osório, F., Crippa, J. A., & Loureiro, S. R. (2007). A study of the discriminative validity of a screening tool (MINI-SPIN) for social anxiety disorder applied to Brazilian university students. *European Psychiatry*, *22*(4), 239-243.

Deary, I. J. (2001*). Intelligence: A very short introduction*. Oxford, UK: Oxford University Press.

Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*, *29*(5), 389-399.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13-21.

Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences*, *98*(17), 10006-10010.

Duchaine, B. C., & Nakayama, K. (2004). Developmental prosopagnosia and the Benton Facial Recognition test. *Neurology*, *62*(7), 1219-1220.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

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

Duesenberg, M., Weber, J., Schulze, L., Schaeuffele, C., Roepke, S., Hellmann-Regen, J., ... & Wingenfeld, K. (2016). Does cortisol modulate emotion recognition and empathy? *Psychoneuroendocrinology, 66*, 221-227.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, *33*(3), 261-268.

Ekman, P. (1992). Are these basic emotions?. In: Ekman, P.; Davidson, R. J. (Eds), *The nature of emotions: Fundamental questions.* Oxford, England: Oxford University Press.

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

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

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

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309-331.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509-529.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Fenske, S., Lis, S., Liebke, L., Niedtfeld, I., Kirsch, P., & Mier, D. (2015). Emotion recognition in borderline personality disorder: effects of emotional information on negative bias. *Borderline Personality Disorder and Emotion Dysregulation*, *2*, 10.

https://doi.org/10.1186/s40479-015-0031-z

Fessler, D. M., Eng, S. J., & Navarrete, C. D. (2005). Elevated disgust sensitivity in the first trimester of pregnancy: Evidence supporting the compensatory prophylaxis hypothesis. *Evolution and Human Behavior*, *26*(4), 344-351.

Fessler, D. M. T., Pillsworth, E. G., & Flamson, T. J. (2004). Angry men and disgusted women: An evolutionary approach to the influence of emotions on risk taking. *Organizational Behavior and Human Decision Processes*, *95*(1), 107–123.

Fett, A. K. J., Viechtbauer, W., Penn, D. L., van Os, J., & Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 573-588.

Fisher, K., Towler, J., & Eimer, M. (2016). Facial identity and facial expression are initially integrated at visual perceptual stages of face processing. *Neuropsychologia*, *80*, 115–125. https://doi.org/10.1016/j.neuropsychologia.2015.11.011

Fleischman, D. S. (2014). Women's disgust adaptations. In V.A. Weekes-Shackelford & T.K. Shackelford (Eds.), *Evolutionary perspectives on human sexual psychology and behavior* (pp. 277-296). New York: Springer.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, *7*(3), 286-299.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2), 171-191.

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. S. (2009). The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *NeuroImage*, *44*(2), 569–580. https://doi.org/10.1016/j.neuroimage.2008.09.011

Fysh, M. C., & Bindemann, M. (2017). The Kent Face Matching Test. *British Journal of*

Psychology, *109*(2), 219–231. https://doi.org/10.1111/bjop.12260

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*(5), 847-857.

Gignac, G. E., Shankaralingam, M., Walker, K., & Kilpatrick, P. (2016). Short-term memory for faces relates to general intelligence moderately. *Intelligence*, *57*, 96-104.

Golan, O., Baron-Cohen, S., Hill, J. J., & Golan, Y. (2006). The "Reading the Mind in Films" Task: Complex emotion recognition in adults with and without autism spectrum conditions. *Social Neuroscience*, *1*(2), 111–123. https://doi.org/10.1080/17470910600980986

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*(2), 93-104.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79-132.

Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Presents*, *9*, 24–29.

Grimshaw, G. M., Bulman-Fleming, M. B., & Ngo, C. (2004). A signal-detection analysis of sex differences in the perception of emotional faces. *Brain and Cognition*, *54*(3), 248-250.

Grynberg, D., Chang, B., Corneille, O., Maurage, P., Vermeulen, N., Berthoz, S., & Luminet, O. (2012). Alexithymia and the processing of emotional facial expressions (EFEs): Systematic review, unanswered questions and further perspectives. *PLoS ONE, 7*(8). http://doi.org/10.1371/journal.pone.0042429

Hagan, C. C., Woods, W., Johnson, S., Calder, A. J., Green, G. G. R., & Young, A. W.

(2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *106*(47), 20010–20015. https://doi.org/10.1073/pnas.0905792106

Hagan, C. C., Woods, W., Johnson, S., Green, G. G. R., & Young, A. W. (2013). Involvement of right STS in audio-visual integration for affective speech demonstrated using MEG. *PLoS One*, *8,* e70648.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, *85*(4), 845-857.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore, MD: Johns Hopkins University Press.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201-206.

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior*, *27*(6), 401-416.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330–337. https://doi.org/10.1016/S1364-6613(00)01519-9

Harmer, C. J., Goodwin, G. M., & Cowen, P. J. (2009). Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *British Journal of Psychiatry*, *195*(2), 102-108.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Heim, A. W. (1970). *AH 4 group test of general intelligence manual.* NFER Publishing

Company Limited.

Herbert, B. M., Herbert, C., & Pollatos, O. (2011). On the relationship between interoceptive awareness and alexithymia: is interoceptive awareness related to emotional awareness?. *Journal of Personality*, *79*(5), 1149-1175.

Hildebrandt, A., Sommer, W., Herzmann, G., & Wilhelm, O. (2010). Structural invariance and age-related performance differences in face cognition. *Psychology and Aging*, *25*(4), 794-810.

Hildebrandt, A., Sommer, W., Schacht, A., & Wilhelm, O. (2015). Perceiving and remembering emotional facial expressions—A basic facet of emotional intelligence. *Intelligence*, *50,* 52-67.

Hildebrandt, A., Wilhelm, O., Schmiedek, F., Herzmann, G., & Sommer, W. (2011). On the specificity of face cognition compared with general cognitive functioning across adult age. *Psychology and Aging, 26*(3), 701-715.

Hobson, H., Brewer, R., Catmur, C., & Bird, G. (2019). The role of language in alexithymia: moving towards a multi-route model of alexithymia. *Emotion Review*. 1-59.

Hoekstra, R. A., Vinkhuyzen, A. A. E., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., … Sluis, S. van der. (2011). The Construction and Validation of an Abridged Version of the Autism-Spectrum Quotient (AQ-Short). *Journal of Autism and Developmental Disorders*, *41*(5), 589–596. https://doi.org/10.1007/s10803-010-1073-0

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta Psychologica*, *135*(3), 278-283.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Articles,* 2.

Horning, S. M., Cornwell, R. E., & Davis, H. P. (2012). The recognition of facial expressions: an investigation of the influence of age and cognition. *Aging, Neuropsychology, and Cognition, 19*(6), 657-676.

Hsing, C. K., Mohr, A. H., Stansfield, R. B., & Preston, S. D. (2013). Alexithymia slows performance but preserves spontaneous semantic decoding of negative expressions in the emostroop task. *International Journal of Psychological Research*, *6*(SPE), 56-67.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424-453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huff, C., & Tingley, D. (2015). “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, *2*(3), 2053168015604648. https://doi.org/10.1177/2053168015604648

Humphreys, K., Minshew, N., Leonard, G. L., & Behrmann, M. (2007). A fine-grained analysis of facial expression processing in high-functioning adults with autism. *Neuropsychologia*, *45*(4), 685-695.

Ioannidis, J. P. A. (2008). Why Most Discovered True Associations Are Inflated. *Epidemiology*, *19*(5), 640–648.

Ioannidis, J. P., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal*, *176*(8), 1091-1096.

Isaacowitz, D. M., & Stanley, J. T. (2011). Bringing an ecological perspective to the study of aging and recognition of emotional facial expressions: Past, current, and future methods. *Journal of Nonverbal Behavior, 35*(4), 261-278.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99,* 561-565.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260-280.

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review*, *3*(4), 371-378.

Jensen, A. R. (1998). *The factor*. Westport CT. Praeger.

Jensen, A. (2001). Vocabulary and general intelligence. *Behavioral and Brain Sciences, 24*(6), 1109-1110. doi:10.1017/S0140525X01280133.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201-211.

Johnson, W., Bouchard Jr, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95-107.

Johnson, W., te Nijenhuis, J., & Bouchard Jr, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence, 36*(1), 81-95.

Joormann, J., & Gotlib, I. H. (2006). Is this happiness I see? Biases in the identification of emotional facial expressions in depression and social phobia. *Journal of Abnormal Psychology*, *115*(4), 705–714. https://doi.org/10.1037/0021-843X.115.4.705

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Mooresville, IL, Scientific Software International.

Kauhanen, J., Kaplan, G. A., Julkunen, J., Wilson, T. W., & Salonen, J. T. (1993). Social factors in alexithymia. *Comprehensive Psychiatry*, *34*(5), 330-335.

Kendler, K. S., Myers, J., Potter, J., & Opalesky, J. (2009). A web-based study of personality, psychopathology and substance use in twin, other relative and relationship pairs. *Twin Research and Human Genetics*, *12*(2), 137-141.

Kerr, S., Johnson, V. K., Gans, S. E., & Krumrine, J. (2004). Predicting adjustment during the transition to college: Alexithymia, perceived stress, and psychological symptoms. *Journal of College Student Development*, *45*(6), 593-611.

Kirouac, G., & Doré, F. Y. (1985). Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behavior*, *9*(1), 3-7.

Klasen, M., Chen, Y. H., and Mathiak, K. (2012). Multisensory emotions: perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*, 381–392. https://doi.org/10.1515/revneuro-2012-0040

Kokkonen, P., Karvonen, J. T., Veijola, J., Läksy, K., & Jokelainen, J. (2001). Perceived and sociodemographic correlates of alexithymia in a population sample of young adults. *Comprehensive Psychiatry*. 471-476.

Korkoliakou, P., Christodoulou, C., Kouris, A., Porichi, E., Efstathiou, V., Kaloudi, E., ... & Douzenis, A. (2014). Alexithymia, anxiety and depression in patients with psoriasis: a case–control study. *Annals of General Psychiatry*, *13*(1), 38.

Kret, M. E., & de Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211-1221.

Krüger, S., Sokolov, A. N., Enck, P., Krägeloh-Mann, I., & Pavlova, M. A. (2013). Emotion through locomotion: gender impact. *PloS One*, *8*(11), e81716.

Kurucz, J., Feldmar, G., & Werner, W. (1979). Prosopo-Affective Agnosia Associated with Chronic Organic Brain Syndrome. *Journal of the American Geriatrics Society*, *27*(2), 91–95. https://doi.org/10.1111/j.1532-5415.1979.tb03347.x

Lane, R., & Schwartz G. E. (1987). Levels of emotional awareness: A cognitive-developmental theory and its application to psychopathology. *American Journal of Psychiatry, 144*, 133-143.

Lane, R. D., Sechrest, L., Reidel, R., Brown, V., Kaszniak, A., & Schwartz, G. (1995). Alexithymia and nonverbal emotion information-processing deficits. *Psychosomatic Medicine, 57*(1), 84.

Lane, R. D., Sechrest, L., Riedel, R., Shapiro, D. E., & Kaszniak, A. W. (2000). Pervasive emotion recognition deficit common to alexithymia and the repressive coping style. *Psychosomatic medicine*, *62*(4), 492-501.

Larsen, J. K., Brand, N., Bermond, B., & Hijman, R. (2003). Cognitive and emotional characteristics of alexithymia: a review of neurobiological studies. *Journal of Psychosomatic Research*, *54*(6), 533-541.

Lau, J. Y., Burt, M., Leibenluft, E., Pine, D. S., Rijsdijk, F., Shiffrin, N., & Eley, T. C. (2009). Individual differences in children's facial expression recognition ability: the role of nature and nurture. *Developmental Neuropsychology*, *34*(1), 37-51.

Lee, N. C., Krabbendam, L., White, T. P., Meeter, M., Banaschewski, T., Barker, G. J., ... & Frouin, V. (2013). Do you see what I see? Sex differences in the discrimination of facial emotions during adolescence. *Emotion*, *13*(6), 1030-1040.

Levin, H. S., Hamsher, K. D. S., & Benton, A. L. (1975). A Short Form of the Test of Facial Recognition for Clinical Use. *The Journal of Psychology*, *91*(2), 223–228. https://doi.org/10.1080/00223980.1975.9923946

Lewis, G. J., Lefevre, C. E., & Young, A. W. (2016). Functional architecture of visual emotion recognition ability: A latent variable approach. *Journal of Experimental Psychology: General*, *145*(5), 589–602. https://doi.org/10.1037/xge0000160

Li, S., Zhang, B., Guo, Y., & Zhang, J. (2015). The association between alexithymia as assessed by the 20-item Toronto Alexithymia Scale and depression: a meta-analysis. *Psychiatry Research*, *227*(1), 1-9.

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: how age shapes emotion processing in nonverbal vocalizations. *Emotion*, *14*(1), 145-160.

Lumley, M. A., Asselin, L. A., & Norman, S. (1997). Alexithymia in chronic pain patients. *Comprehensive Psychiatry*, *38*(3), 160-165.

Lumley, M. A., Gustavson, B. J., Partridge, R. T., & Labouvie-Vief, G. (2005). Assessing alexithymia and related emotional ability constructs using multiple methods: Interrelationships among measures. *Emotion*, *5*(3), 329-342.

Lumley, M. A., Neely, L. C., & Burger, A. J. (2007). The assessment of alexithymia in medical settings: implications for understanding and treating health problems. *Journal of Personality Assessment*, *89*(3), 230-246.

Lyusin, D., & Ovsyannikova, V. (2016). Measuring two aspects of emotion recognition ability: Accuracy vs. sensitivity. *Learning and Individual Differences, 52,* 129-136.

Lyvers, M., Onuoha, R., Thorberg, F. A., & Samios, C. (2012). Alexithymia in relation to parental alcoholism, everyday frontal lobe functioning and alcohol consumption in a non-clinical sample. *Addictive Behaviors*, *37*(2), 205-210.

Mansell, W., Clark, D. M., Ehlers, A., & Chen, Y. P. (1999). Social anxiety and attention away from emotional faces. *Cognition & Emotion*, *13*(6), 673-690.

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology*, *626*(1), 83-86.

Martens, U., Leuthold, H., & Schweinberger, S. R. (2010). Parallel processing in face perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 103-121.

Matsumoto, D., & Hwang, H. S. (2011). Judgments of facial expressions of emotion in profile. *Emotion*, *11*(5), 1223-1229.

Mattila, A. K., Salminen, J. K., Nummi, T., & Joukamaa, M. (2006). Age is strongly associated with alexithymia in the general population. *Journal of Psychosomatic Research*, *61*(5), 629-635.

Maylor, E. A., Moulson, J. M., Muncer, A. M., & Taylor, L. A. (2002). Does performance on theory of mind tasks decline in old age?. *British Journal of Psychology, 93*(4), 465-485.

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(21), 1-15. https://doi.org/10.1186/s41235-018-0112-9

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424-453.

Meganck, R., Vanheule, S., & Desmet, M. (2008). Factorial validity and measurement invariance of the 20-item Toronto Alexithymia Scale in clinical and nonclinical samples. *Assessment*, *15*(1), 36-47.

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion*, *9*(5), 619-630.

Miller, J. D., Crowe, M., Weiss, B., Maples-Keller, J. L., & Lynam, D. R. (2017). Using online, crowdsourcing platforms for data collection in personality disorder research: The example of Amazon's Mechanical Turk. *Personality Disorders: Theory, Research, and Treatment*, *8*(1), 26.

Mogg, K., Philippot, P., & Bradley, B. P. (2004). Selective attention to angry faces in clinical social phobia. Journal of abnormal psychology, 113(1), 160-165.

Montebarocci, O., Surcinelli, P., Rossi, N., & Baldaro, B. (2011). Alexithymia, verbal ability and emotion recognition. *Psychiatric Quarterly*, *82*(3), 245-252.

Mooney, C. M. (1956). Closure with negative after-images under flickering light. *Canadian Journal of Psychology, 10*(4), 191–199.

Mortensen, K., & Hughes, T. L. (2018). Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. *Journal of General Internal Medicine, 33*(4), 533-538.

Murphy, J., Catmur, C., & Bird, G. (2018). Alexithymia is associated with a multidomain, multidimensional failure of interoception: Evidence from novel tests. *Journal of Experimental Psychology: General*, *147*(3), 398-408.

Murphy, N. A., & Isaacowitz, D. M. (2010). Age effects and gaze patterns in recognising emotional expressions: An in-depth look at gaze measures and covariates. *Cognition and Emotion*, *24*(3), 436-452.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, *51*(2), 77-101.

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, *1*(3), 261-289.

Newcombe, F. (1979). The processing of visual information in prosopagnosia and acquired dyslexia: functional versus physiological interpretation. In D. J. Oborne, M. M. Gruneberg, & J. R. Eiser (Eds.), *Research in psychology and medicine, Vol 1* (pp. 315-322). London: Academic Press.

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior, 18*(1), 9-35.

Olderbak, S., Wilhelm, O., Hildebrandt, A., & Quoidbach, J. (2018). Sex differences in facial emotion perception ability across the lifespan. *Cognition and Emotion*, 1-10.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), doi: 10.1126/science.aac4716.

Ornaghi, V., & Grazzani, I. (2013). The relationship between emotional-state language and emotion understanding: A study with school-age children. *Cognition & Emotion*, *27*(2), 356-366.

Ozonoff, S., Heung, K., Byrd, R., Hansen, R., & Hertz-Picciotto, I. (2008). The onset of autism: patterns of symptom emergence in the first years of life. *Autism Research*, *1*(6), 320-328.

Palermo, R., O'Connor, K. B., Davis, J. M., Irons, J., & McKone, E. (2013). New Tests to Measure Individual Differences in Matching and Labelling Facial Expressions of Emotion, and Their Association with Ability to Recognise Vocal Emotions and Facial Identity. *PLOS ONE*, *8*(6), e68126. https://doi.org/10.1371/journal.pone.0068126

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411-419.

Parker, J. D., Taylor, G. J., & Bagby, R. M. (1989). The alexithymia construct: relationship with sociodemographic variables and intelligence. *Comprehensive Psychiatry*, *30*(5), 434-441.

Parry, F. M., Young, A. W., Shona, J., Saul, M., & Moss, A. (1991). Dissociable face processing impairments after brain injury. *Journal of Clinical and Experimental Neuropsychology*, *13*(4), 545–558. https://doi.org/10.1080/01688639108401070

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal Representations of Perceived Emotions in the Human Brain. *Journal of Neuroscience*, *30*(30), 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organisation and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London, Series B., 335*, 23-30.

Peterson, E., & Miller, S. (2012). The eyes test as a measure of individual differences: how much of the variance reflects verbal IQ?. *Frontiers in psychology, 3*, 220. https://doi.org/10.3389/fpsyg.2012.00220

Preece, D., Becerra, R., Allan, A., Robinson, K., & Dandy, J. (2017). Establishing the theoretical components of alexithymia via factor analysis: Introduction and validation of the attention-appraisal model of alexithymia. *Personality and Individual Differences*, *119*, 341-352.

Preece, D. A., Becerra, R., Boyes, M. E., Northcott, C., McGillivray, L., & Hasking, P. A. (2020). Do self-report measures of alexithymia measure alexithymia or general psychological distress? A factor analytic examination across five samples. *Personality and Individual Differences*, *155*, 109721.

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and Cognition*, *54*(3), 179-185.

Ramsey, S. R., Thompson, K. L., McKenzie, M., & Rosenbaum, A. (2016). Psychological research in the internet age: The quality of web-based data. *Computers in Human Behavior, 58*, 354-360.

Rao, P. A., Beidel, D. C., & Murray, M. J. (2008). Social skills interventions for children with Asperger’s syndrome or high-functioning autism: A review and recommendations. *Journal of Autism and Developmental Disorders*, *38*(2), 353-361.

Rapcsak, S. Z., Galper, S. R., Comer, J. F., Reminger, S. L., Nielsen, L., Kaszniak, A. W., ... & Cohen, R. A. (2000). Fear recognition deficits after focal brain damage: a cautionary note. *Neurology*, *54*(3), 575-598.

Rhodes, G., Pond, S., Burton, N., Kloth, N., Jeffery, L., Bell, J., … Palermo, R. (2015). How distinct is the coding of face identity and expression? Evidence for some common dimensions in face space. *Cognition*, *142*, 123–137. https://doi.org/10.1016/j.cognition.2015.05.012

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029-2040.

Rosenberg, H., Dethier, M., Kessels, R. P., Westbrook, R. F., & McDonald, S. (2015). Emotion perception after moderate–severe traumatic brain injury: The valence effect and the role of working memory, processing speed, and nonverbal reasoning. *Neuropsychology*, *29*(4), 509-521.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*(2), 1-36.

Rossion, B. (2018). Normative accuracy and response time data for the computerized Benton Facial Recognition Test (BFRT-c). *Behavior Research Methods*, *50*, 2442-2460.

Rossion, B., & Michel, C. (2018). Normative accuracy and response time data for the computerized Benton Facial Recognition Test (BFRT-c). *Behavior Research Methods*, *50*(6), 2442-2460.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*(2), 139-148.

Rozin, P., Haidt, J., & Fincher, K. (2009). From Oral to Moral. *Science*, *323*(5918), 1179–1180. https://doi.org/10.1126/science.1170492

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology, 66*(5), 870-881.

Rozin, P., Taylor, C., Ross, L., Bennett, G., & Hejmadi, A. (2005). General and specific abilities to recognise negative emotions, especially disgust, as portrayed in the face and the body. *Cognition & Emotion*, *19*(3), 397-412.

Ruffman, T. (2011). Ecological validity and age-related change in emotion recognition. *Journal of Nonverbal Behavior*, *35*(4), 297-304.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, *32*(4), 863-881.

Rutter, L. A., Dodell-Feder, D., Vahia, I. V., Forester, B. P., Ressler, K. J., Wilmer, J. B., & Germine, L. (2019). Emotion sensitivity across the lifespan: Mapping clinical risk periods to sensitivity to facial emotion intensity. *Journal of experimental psychology. General.* 1-13.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism, 6*(1), 2.

Sacco, D. F., & Hugenberg, K. (2009). The look of fear and anger: Facial maturity modulates recognition of fearful and angry expressions. *Emotion*, *9*(1), 39-49.

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International neuropsychological Society*, *16*(5), 754-760.

Salthouse, T. A. (2013). Within-cohort age-related differences in cognitive functioning. *Psychological Science*, *24*(2), 123-130.

Salthouse, T. A. (2014a). Why are there different age relations in cross-sectional and longitudinal comparisons of cognitive functioning?. *Current Directions in Psychological Science*, *23*(4), 252-256.

Salthouse, T. A. (2014b). Aging cognition unconfounded by prior test experience. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 71*(1), 49-58.

Sasson, N. J., Pinkham, A. E., Richard, J., Hughett, P., Gur, R. E., & Gur, R. C. (2010). Controlling for response biases clarifies sex and age differences in facial affect recognition. *Journal of Nonverbal Behavior*, *34*(4), 207-221.

Sawada, R., Sato, W., Kochiyama, T., Uono, S., Kubota, Y., Yoshimura, S., & Toichi, M. (2014). Sex differences in the rapid detection of emotional facial expressions. *PloS One*, *9*(4), e94747.

Schirmer, A., & Adolphs, R. (2017). Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence. *Trends in Cognitive Sciences*, *21*(3), 216–228. https://doi.org/10.1016/j.tics.2017.01.001.

Schlegel, K., & Scherer, K. R. (2016). Introducing a short version of the Geneva Emotion Recognition Test (GERT-S): Psychometric properties and construct validation. *Behavior Research Methods, 48*(4), 1383-1392.

Schlegel, K., Grandjean, D., & Scherer, K. R. (2012). Emotion recognition: Unidimensional ability or a set of modality- and emotion-specific skills? *Personality and Individual Differences*, *53*(1), 16–21. https://doi.org/10.1016/j.paid.2012.01.026

Schretlen, D. J., Pearlson, G. D., Anthony, J. C., & Yates, K. O. (2001). Determinants of Benton Facial Recognition Test performance in normal adults. *Neuropsychology*, *15*(3), 405-410.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40*(1-2), 99-116.

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1748-1765.

Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition*, *27*(3), 538-552.

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., … Matthews, F. E. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*, 204. https://doi.org/10.1186/s12883-014-0204-1

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences*, *112*(41), 12887–12892. https://doi.org/10.1073/pnas.1421881112

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, *44*(3), 535-551.

Shuttleworth, E. C., Syring, V., & Allen, N. (1982). Further observations on the nature of prosopagnosia. *Brain and Cognition*, *1*(3), 307–322. https://doi.org/10.1016/0278-2626(82)90031-8

Smith, T., Marsden, P. V., Hout, M., & Kim, J. (2012). *General Social Surveys, 1972–2012 Cumulative Codebook.* Chicago: National Opinion Research Centre.

Sogon, S., & Izard, C. E. (1987). Sex differences in emotion recognition by observing body movements: A case of American students. *Japanese Psychological Research*. *29*(2), 89-93.

Sokolov, A. A., Krüger, S., Enck, P., Krägeloh-Mann, I., & Pavlova, M. A. (2011). Gender affects body language reading. *Frontiers in Psychology*, *2*(16), 1-6.

Spearman, C. (1927). *The abilities of man* (Vol. 89). New York: Macmillan.

Sprengelmeyer, R., Young, A. W., Pundt, I., Sprengelmeyer, A., Calder, A. J., Berrios, G., ... & Przuntek, H. (1997). Disgust implicated in obsessive–compulsive disorder. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1389), 1767-1773.

Stevenson, R. J., Oaten, M. J., Case, T. I., Repacholi, B. M., & Wagland, P. (2010). Children's response to adult disgust elicitors: Development and acquisition. *Developmental Psychology, 46*(1), 165-177.

Surguladze, S. A., Young, A. W., Senior, C., Brébion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition accuracy and response bias to happy and sad facial expressions in patients with major depression. *Neuropsychology, 18*(2), 212-218.

Suzuki, A., Hoshino, T., & Shigemasu, K. (2010). Happiness is unique: A latent structure of emotion recognition traits revealed by statistical model comparison. *Personality and Individual Differences*, *48*(2), 196-201.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, *28*(7), 1164-1195.

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*(4), 42–50. https://doi.org/10.1109/38.946630

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*(4), 578–598. https://doi.org/10.1037/0021-843X.112.4.578

Tombaugh, T. N. (2006). A comprehensive review of the paced auditory serial addition test

(PASAT). *Archives of Clinical Neuropsychology*, *21*(1), 53-76.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science*, *280*(5365), 867-873.

Ventura, J., Wood, R. C., Jimenez, A. M., & Hellemann, G. S. (2013). Neurocognition and symptoms identify links between facial recognition and emotion processing in schizophrenia: Meta-analytic findings. *Schizophrenia Research*, *151*(1), 78–84. https://doi.org/10.1016/j.schres.2013.10.015

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Bargary, G., Lawrance-Owen, A. J., & Mollon, J. D. (2014). An online version of the Mooney Face Test: phenotypic and genetic associations. *Neuropsychologia*, *63*, 19-25. http://dx.doi.org/10.1016/j.neuropsychologia.2014.08.011

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217–227. https://doi.org/10.1016/j.visres.2016.12.014

Walk, R. D., & Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bulletin of the Psychonomic Society*, *22*(5), 437-440.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.

Weiss, M. J., & Harris, S. L. (2001). Teaching social skills to people with autism. *Behavior Modification*, *25*(5), 785-802.

West, J. T., Horning, S. M., Klebe, K. J., Foster, S. M., Cornwell, R. E., Perrett, D., ... & Davis, H. P. (2012). Age effects on emotion recognition in facial displays: From 20 to 89 years of age. *Experimental Aging Research*, *38*(2), 146-168.

White, S. W., Bray, B. C., & Ollendick, T. H. (2012). Examining Shared and Unique Aspects of Social Anxiety Disorder and Autism Spectrum Disorder Using Factor Analysis.

Journal of Autism and Developmental Disorders, 42(5), 874–884. https://doi.org/10.1007/s10803-011-1325-7.

Wieck, C., & Kunzmann, U. (2017). Age differences in emotion recognition: A question of modality?. *Psychology and Aging, 32*(5), 401-411.

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces—One element of social cognition. *Journal of Personality and Social Psychology, 99*(3), 530-548.

Williams, M. A., & Mattingley, J. B. (2006). Do angry men get noticed? *Current Biology*, *16*(11), 402-404.

Wilmer, J. B., Germine, L. T., & Nakayama, K. (2014). Face recognition: a model specific ability. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00769

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., … Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107.

Wu, W., West, S. G., & Taylor, A. B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods*, *14*(3), 183-201.

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research*, *65*(5), 445-451.

Young, A. (2016). *Facial expression recognition*. Hove, East Sussex: Psychology Press.

Young, A. W. (2018). Faces, people and the brain: The 45th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology, 71*(3), 569–594.

https://doi.org/10.1177/1747021817740275

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology*, *102*(4), 959-974.

Young, A. W., & Burton, A. M. (2017). Recognizing Faces. *Current Directions in Psychological Science*, *26*(3), 212–217. https://doi.org/10.1177/0963721416688114

Young, A. W., & Burton, A. M. (2018). Are we face experts? *Trends in Cognitive Sciences, 22*, 100-110.

Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, *116*(4), 941–959. https://doi.org/10.1093/brain/116.4.941

Young, A. W., Perrett, D., Calder, A., Sprengelmeyer, R., & Ekman, P. (2002). *Facial expressions of emotion – stimuli and tests (FEEST)*. Bury St Edmunds, England: Thames Valley Test Company.

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

Zhang, W., & Luck, S. J. (2008). Discrete Fixed-Resolution Representations in Visual Working Memory. *Nature*, *453*(7192), 233–235.