DOCTOR OF PHILOSOPHY

Facial cues to mental health symptoms

Scott, Naomi

Award date:
2015

Awarding institution:
Bangor University

Facial Cues to Mental Health Symptoms

Naomi Scott

School of Psychology

Bangor University

Bangor, UK

• PRIFYSGOL CYMRU •
UNIVERSITY OF WALES
BANGOR
1884

This thesis is submitted to the School of Psychology, Bangor University, Wales, in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Acknowledgements

Firstly to Tash, thank-you for making me laugh every day, for encouraging me to try life outside of my comfort zone and putting everything into perspective. Your love and support has encouraged me to be confident in who I am and what I can do. You have inspired me to reach for my dreams, and supported them no matter how ludicrous. I can’t wait for the next few years and the adventures we will have along the way.

A massive thank-you to my supervisor, Rob Ward, for encouraging my passion for this work during my Masters, and giving me the opportunity to develop it over the last three years. I’d also like to thank Alex and Robin, whose early encouragement and support in the cruel mistress’ of MatLab and R allowed me to develop this thesis as far as I have.

To Mandy and Sofia, who shared this experience with me, I couldn’t have wished for two more amazing individuals to share the journey with. I will be eternally grateful for the unwavering support, lunchtime escapes and the endless, slightly questionable, humour. I’d also like to thank the F30 girls, who enticed me away from the office with roast dinners, opportunities for questionable dancing and an unhealthy obsession with Christmas movies, which made for an epic couple of years. I feel privileged to have found five lifelong friends in you all.

Not forgetting some of my oldest, and dearest friends. To Laura, without your encouragement and support through my Masters year I would not have had the confidence to even apply for this funding and to Ailsa, who has been there through thick and thin, I’m sorry I couldn’t make it home as often as we’d like but one day we’ll be in the same country!

I’d like to dedicate this thesis to my Dad, who installed scientific curiosity and a love of numbers in me from a young age. I’m sure if he knew about this thesis he would have been silently proud, whilst making some joke about a lifelong student and not getting a ‘proper’ job.

Contents

Thesis Summary

This thesis comprises of seven experimental chapters which demonstrate the ability for naïve observers to accurately distinguish between facial stimuli of individuals scoring high on measures of neuropsychiatric disorders such as depression, borderline and schizotypal personality disorders and autistic spectrum disorder. In the case of depression this ability was still apparent even when the stimuli was reduced to show just the eye and brow region. The findings that observers are able to accurately discriminate between this stimuli furthers the vast literature of face research, demonstrating that this simple stimuli can be used to discriminate traits previously only assessed using thin-slice video stimuli.

In addition to this within each study observers were asked to make inferences about socially desirable personality traits, assessed using the big five personality traits. A consistent, negative, personality-type was attributed to individuals with high symptoms levels - commonly consisting of high levels of Neuroticism combined with low levels of Agreeableness and Extraversion. This downgrading of desirable personality traits for individuals scoring highly for neuropsychiatric traits has potential social implications for individuals with these disorders. The traits these individuals are seen to be lower in have been shown to be important in the development and maintenance of successful social relationships, thus if individuals with neuropsychiatric disorders are perceived to be lower in these traits there is the potential that they are at a disadvantage during social interactions.

The final section of the thesis looks to link these findings to existing theories of neuropsychiatric disorders. Results demonstrate a masculine facial appearance associated with males with symptoms of autism, supporting Baron-Cohen's (2002) masculinity hypothesis, and findings that individuals with eating disorders and

depression are perceived to be more feminine lend support to Badcock and Crespi's (2008) diametric hypothesis. These studies highlight a perceptual continuum of gender underlying observations of individuals with neuropsychiatric disorders, where autism lies at the masculine end with attributions of positive personality traits and psychotic disorders are at the feminine end with attributions of negative personality traits. These findings can be associated with Todorov, Said, Engell and Oosterhof's (2008) theory that perceptions of facial stimuli lie on a two dimensional continuum of trustworthiness and dominance (including assumptions of masculinity). This suggests that perceptions of neuropsychiatric disorders may be attributable to an underlying continuum of social desirability, reflecting the findings of a downgrading of personality traits to high scoring individuals.

This thesis extends current facial research, demonstrating that a number of common neuropsychiatric disorders can be distinguished from very simple facial stimuli. Not only this, but a distinct pattern of negative personality traits are attributed to individuals with high levels of neuropsychiatric traits. These combined with the perceptions of gender and the stereotypes associated with these perceptions have potential implications for social interaction. That is, if individuals with neuropsychiatric traits are perceived to have personality traits correlated with negative social connotations, others may be less likely to interact with them, reinforcing some of the symptoms of these disorders.

Chapter one: Introduction

"The face is a picture of the mind with the eyes as its interpreter"

Marcus Tullius Cicero

Recent predictions estimate that 6.2% of the adult UK population meets criteria for at least one of the six common mental health illnesses (anxiety, mixed anxiety and depression, depressive episode, phobias, obsessive compulsive disorder and panic disorder; Office of National Statistics, 2007). The most prevalent of these disorders, mixed anxiety and depressive disorder, is predicted to be the 2nd largest cause of disability in developed countries by 2030 (Mathers & Loncar, 2006). As expected, these illnesses have a large economic impact, with a UK estimated economic and social cost of £105.2 billion in 2009-2010 (Sailsbury Centre, 2010). Not only do they have a widespread economic cost, but they also have social costs at an individual level; for example, those with depressive disorders are more likely to have no educational qualifications and be in a lower social class and less likely to be married or cohabiting (Singleton, Bumpstead, O'Brien, Lee, & Meltzer, 2001).

The prevalence of these disorders is well documented, however the success rate of early diagnosis and treatment is less successful. Although Singleton and colleagues (2001) found 15% of surveyed individuals reported a degree of neurotic symptoms worthy of treatment, less than a quarter of them were actually receiving any treatment, even though 39% reported speaking to their GP about a mental or emotional problem they were experiencing.

Currently the diagnosis and treatment of mental illnesses relies on the individual presenting symptoms to healthcare professionals, but are there additional cues that could be utilised to aid the early diagnosis of psychiatric disorders? Mehl (2006) identified that non-professional participants could accurately identify depression from activity diaries. Participants were asked to give acoustic logs of social encounters for two days, which were then viewed by naïve judges who rated the participants level of depression. High BDI scorers were rated as more depressed than controls, although a general correlation across the full range of BDI scorers was not found. Judges were more likely to rate participants as depressed if they spent time alone, were not talkative and sought out little entertainment or socialising. These cues match the experiences of men with depression, identified in a qualitative study by Heifner (1997), such as feelings of a lack of connectedness and integration with others.

Thin slice cues to psychiatric disorders

Although these studies suggest there are behavioural cues to psychiatric disorders such as depression they require in-depth analysis, which require time to both collect and analyse. However, thin slice studies have explored the possibility that we are able accurately identify a range of traits from very limited exposure to individual’s behaviour, utilising random samples of behaviour lasting less than 5 minutes in length, either by full video stream or silent videos (Ambady, 2010). Indeed this concept is not new, on reviewing video footage of a suicidal patient claiming to have recovered from depression whilst requesting a weekend pass from a psychiatric hospital, Ekman (1985) noted a “fleeting look of despair” before responding to a question about her plans for the future. This non-verbal cue was often missed by clinicians, however it suggests that viewing thin slices of behaviour can highlight clinically relevant traits.

In a comprehensive review, Slepian, Bogart and Ambady (2014) discuss the findings of how thin slice studies can be utilised in a clinical context. Here, Slepian and colleagues present evidence for the accurate recognition of personality disorders, anxiety and depression utilising a variety of methodologies, the findings of which I shall review in turn.

Personality disorders: The accurate identification of the personality traits Extraversion and Openness has been demonstrated using thin slice studies when observers are presented with films both including and missing sound (Borkenau & Liebler, 1993) and when segments were reduced to just 5 seconds in length (Carney, Colvin, & Hall, 2007). This ability was extended to all of the Big Five personality traits (Agreeableness, Conscientiousness, Extraversion, Openness and Neuroticism) when the stimulus was extended to 20-seconds in length (Carney et al., 2007). Slepian and colleagues (2014) propose that, as personality disorders are considered maladaptive variants of personality traits, which can be accurately identified from thin slices of behaviour, personality disorders may be observable too. Research suggests that this is indeed the case, with observers being able to accurately identify personality traits associated with schizoid, avoidant, borderline, antisocial and obsessive compulsive personality disorders (Oltmanns, Friedman, Fiedler, & Turkheimer, 2004). These findings were furthered by Friedman, Oltmanns and Turkheimer (2007) showing a significant correlation between self and observer ratings of schizoid, histronic, narcisstic and avoidant personality disorders and ratings from observers after viewing 30-second video clips. The finding that these results also correlated to friend's ratings of the participant's personality disorders strengthens the conclusion that cues to personality disorders can be utilised from thin slice stimuli. However, it must be noted that there was little specificity across personality disorder clusters, which the authors attribute to considerable overlap found between the disorders suggesting a commonality in perceptions over disorders.

This style of study has also been utilised to investigate the ability to detect psychiatric traits, for example Fowler, Lilienfeld and Partick (2009) showed that observers were able to accurately identify psychopathy from 5-second video clips of maximum security inmates. Utilising a dyadic scenario methodology Rauthmann (2011) investigated how individuals with dark personality disorders (narcissism, psychopathy and Machiavellianism) are viewed by others during initial, brief, interactions. Narcissists were perceived to be arrogant, machiavellians were perceived to be low in dominance, openness and sociableness and psychopaths to be highly dominant and arrogant, and low in nurturance and conscientiousness. Furthermore, Vazire, Naumann, Rentfrow and Gosling (2008) determined that observers could accurately identify narcissism from full body images. Cues utilised to make this attribution included expensive and stylish clothes, a neat appearance and attractiveness.

Anxiety: In a meta-analysis of 80 studies assessing auditory and visual cues to trait and state anxiety Harrigan, Wilson and Rosenthan (2004) concluded that observers were able to utilise cues to accurately identify both disorders, although the channel for optimal accuracy differed. For state anxiety the overall effect size was higher when audio-only stimuli was utilised, for trait anxiety video-only stimuli was more accurate.

Depression: The accurate identification of depression has been demonstrated using even less information, with participants ratings significantly correlating with patients Minnesota Multiphasic Personality Inventory scores after viewing 20-seconds of silent videotape (Waxer, 1976). In a similar study Carney and colleagues (2007) demonstrated above chance level detection of negative affect from as little as 5-seconds exposure to videoed dyadic interactions, and positive affect from 20-second exposure. Interestingly in this study, accuracy was greater when participants were shown excerpts from the 3rd or 5th minute of these interactions, suggesting the middle or end of an interaction holds more reliable cues to these traits.

In addition to these studies Slepian and colleagues (2014) discuss how thin slice studies can be utilised to predict long term wellbeing. For example, Mason, Sbarra and Mehl (2010) found observers able to accurately predict both current and future self-reported psychological adjustment after a recent divorce from 30-second sound clips of the individuals talking about their ex-partners. Interestingly, accuracy was not maintained when observers read transcripts of the same clips, suggesting the way participants spoke that held the cues rather than the content.

The studies discussed above show how short excerpts of behaviour can be utilised to develop accurate impressions of psychiatric traits, even when stimuli is reduced to 5-seconds in length or impoverished by the removal of sound. In a meta-analysis of thin slice studies Ambady (2010) concluded that the judgements used here are efficient, made intuitively and are in fact more accurate when made without deliberation. However, these clips still hold a relatively large amount of information, and accuracy for traits has been demonstrated when the stimuli is reduced even further. For example Ambady, Hallahan and Conner (1999) demonstrated the ability to correctly categorise males as heterosexual or homosexual when presented with 10-second and 1-second silent video clips as well as when simply a still picture was presented. Utilising a clinical population Grossman (2014) found children with high functioning Autism were rated as more socially awkward than typically developing controls regardless of stimuli type (full video, audio only, visual only or still images).

Indeed perceptions of traits have been shown to hold their accuracy even when still facial images are presented for a limited time. Stillman, Maner and Bausneister (2010) asked participants to rate facial images of violent and non-violent sex offenders for the likelihood that they would display violent behaviour. After presenting the facial images for just two seconds participants rated the violent offenders to be significantly more likely to be violent than the non-violent offenders. Todorov, Pakrashi and Oosterhod (2009) reduced the length that stimuli was presented even further, demonstrating participants ability to accurately detect trustworthiness when faces were presented for just 33 milliseconds, however accuracy did not peak until 100 milliseconds exposure.

These studies demonstrate that a variety of social traits can be accurately inferred from viewing static facial images, which is perhaps unsurprising when we consider that the face holds information about social intentions and motivations (Schmidt & Cohn, 2001) and is one of the first things we see during social interactions. Indeed we appear to be pre-programmed to respond to facial stimuli - with newborn babies paying more attention to faces than any other stimuli (Goren, Sarty, & Wu, 1975). The use of judgements made from faces is suggested to be unintentional and made without awareness (Hassin & Toupe, 2000, Rule, Garrett, & Ambady, 2010) are sustained over time (Sunnafrank & Ramirez, 2004), even when additional information is given (Todorov, Madisodza, Goren, & Hall, 2005). Interestingly Rule, Ambady and Hallett (2009) found that inferences made from static facial images were more accurate when they were made using a ‘snap’ judgement than when participants deliberated over their decision. Torodov and colleagues (2009) found that, even though accuracy did not increase as exposure time did, participants became more confident in their decisions as time increased.

Thin slice studies utilising facial images, as discussed thus far, have generally obtained stimuli by taking still clips from videoed behaviour, which has the potential to include additional information such as clothing, body posture and facial expression.

More recent studies have reduced these biases and shown the accurate detection of a vast range of social traits from static facial photographs with neutral facial poses, such as religious beliefs (Rule et al., 2010), aggression (Carre, McCormick, & Mondloch, 2009), dominance (Meuller & Mazur, 1997; Quist, Watkins, Smith, DeBruine, & Jones, 2010), socio-sexuality (Boothroyd, Jones, Burt, DeBruine, & Perrett, 2008), female (Rule et al., 2009) and male homosexuality (Rule & Ambady, 2008a), criminality (Valla, Ceci, & Williams, 2011), health (Little, McPherson, Dennington, & Jones, 2011), strength (Fink, Neave, & Seydel, 2007), success (Rule & Ambady, 2008b, 2010a) and political affiliation (Rule & Ambady, 2010b).

The ability to detect traits from static facial images is not culturally bound, as Sell et al. (2009) showed that western participants were able to accurately identify strength from viewing facial images of members of a Bolivian tribe. Additionally there is evidence to suggest that some ratings may be consistent over time, with participants demonstrating the ability to predict later career success from yearbook photos (Rule & Ambady, 2011).

One field of research which has received consistent attention in recent years is the ability to detect personality traits from static facial images, coinciding with Hassin and Trope's (2000) findings that 75% of Israeli's believe you can determine an individuals true personality from looking at their face. These findings further the thin-slice studies into the recognition of the Big Five (Agreeableness, Conscientiousness, Extraversion, Openness and Neuroticism) personality traits I discussed earlier (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; Carney et al., 2007). One of the first studies into this field, conducted by Shevlin, Walker, Davies, Banyard and Lewis (2003), asked participants to rate photographs of individuals for Psychoticism, Extraversion and Neuroticism. Participants were shown to be accurate in their ratings of Psychoticism when viewing faces cropped down to show just the face, however results for Extraversion and Neuroticism failed to rise above chance.

This was followed by a two-part study by Penton-Voak, Pound, Little and Perrett (2006). Firstly, participants were presented with individual faces and asked to rate them on each of the big five personality traits. These ratings were then correlated with self-ratings provided by the photographed individuals. Significant correlations were found for ratings of Extraversion, Openness and Neuroticism in male images, but only Extraversion in female images. In study two the photographs of the 15 highest and lowest scoring individuals in each trait were averaged together to create a 'composite' image. This 'morphing' procedure aims to extract the subtle facial characteristics which define a group whilst removing those which make each face individual (Rowland & Perrett, 1995). It is aimed that using these composite images will remove individual variations which are not associated with the personality attributions in question, as well as temporary characteristics such as hair style. When participants were presented with these stimuli by Penton-Voak et al. (2006) they were able to accurately identify Agreeableness and Extraversion in both sexes, as well as Neuroticism in males.

Again using composite facial images Little and Perrett (2007) showed that stimuli created from females scoring higher on the personality traits Agreeableness, Conscientiousness, Extraversion and Neuroticism were rated to be significantly higher in each trait than images created from females scoring lowly. However, only results for Extraversion reached significance for male composites. Once more using composite facial images Kramer and Ward (2010) replicated previous findings, demonstrating accurate identification of Agreeableness, Extraversion and Neuroticism in both sexes. Following these findings Kramer and Ward (2010) further reduced the stimuli, presenting participants with 'T' shaped facial stimuli only showing the internal facial features (eyes, nose and mouth). Interestingly participants were still able to discriminate between high and low composites for each of the previously identified traits, as well as a newfound accuracy for the trait openness. Although these studies control for more additional cues than thin slice studies, they do not take into account contributions of viewing angle or posture. This bias was assessed by Jones, Kramer and Ward (2012) who used three-dimensional images to control for postural cues, finding the accurate identification of Agreeableness and Neuroticism. These results suggest postural cues, such a slight head tilt, may be important in the accurate identification of Extraversion, as previous studies (Kramer & Ward, 2010; Penton-Voak et al., 2006) demonstrated above chance discrimination with standard composite images. In addition, Jones et al. identified a laterality bias, with participants exhibiting an increased accuracy when viewing the right side of the face than the left.

These studies have demonstrated the ability to detect personality traits from static facial images, but how does this link to psychiatric disorders?

The role of Neuroticism in Psychiatric disorders

One of the Big Five personality traits, Neuroticism (characterised by a tendency for low moods, feelings of guilt, envy, anger and anxiety) has been consistently linked to an increased likelihood of developing Neuropsychiatric disorders. For example individuals with high Neuroticism scores are more likely to experience distress (Bagby et al., 1996), which can go on to develop into depressive symptoms (Bagby et al., 1997). Additional support comes from the findings that individuals with depression have consistently higher than average levels of Neuroticism (Bagby et al., 1997; Trull & Sher, 1994; Widiger & Trull, 1992) and that Neuroticism has been utilised as a cue for later depressive symptoms (Roberts & Kendler, 1999; Van Os, Park, & Jones, 2001). Indeed Neuroticism has been proposed to be a diathesis for depression (Krueger, Caspi, & Moffitt, 2000; Trull & Sher, 1994). These associations are more widespread than just depressive symptoms, for example Muntaner, Garcia-Sevilla, Fernandez and Torrubia (1988) found both schizotypal and borderline personality disorders to be positively associated with Neuroticism. Similarly Van Os and Jones (2001) found a significant, positive, association between Neuroticism and Schizophrenia, as well as a negative association with Extraversion. Wakabayashi, Baron-Cohen and Wheelwright (2006) extend these findings to autistic-spectrum disorder, demonstrating a positive correlation between Autistic-Spectrum Quotient scores and Neuroticism as well as a negative correlation with Conscientiousness and Extraversion.

Two or five dimensions of personality?

Although the Big Five personality traits are widely used within social psychology another theory may underlie the apparent accurate perception of personality traits from facial features. This two-dimensional approach to personality was first highlighted by Asch's (1946) studies into person perception, contrasting a competent person who was warm to a competent person who was cold. This has been followed by a wealth of research investigating the possibility that social perceptions of others can be boiled down to two orthogonal concepts. Fisk, Cuddy and Glick (2007) define these two concepts as warmth and competence, where warmth captures traits related to perceived intent, including friendliness, helpfulness, trustworthiness, sincerity and morality and competence, capturing traits of perceived ability, such as intelligence, skill, creativity and efficacy. Wojciszke, Dowhyluk and Jaworski (1998) lent support for the notion that these two concepts explain how people characterise others, demonstrating that they account for 82% of variance in every day behaviour. In addition, Wojciszke (1994) asked participants to recount real-life encounters and found participants framed in excess of 75% of over 1000 events in terms of morality (warmth) and competence.

Fisk and colleagues (2007) suggests that the social categorisation of individuals in this manner is a consequence of the evolutionary concept of fight or flight. Warmth is used as an indicator to determine if a person has positive or negative intentions, and competency a signal as to whether they are able to carry out those intentions. The proposal that this is an evolved ability is supported by White's (1982) findings that the concepts are stable across cultures. Findings that words related to warmth are identified quicker than those relating to competence (Ybarra, Chen & Park, 2001) suggests that warmth is the primary, and competence the secondary judgement made. This creates a two-step approach with warmth judgments carrying more weight in affective and behavioural interactions, whereas the second judgement, competence, is a more diagnostic perception of the other persons abilities.

Judd, James-Hawkins, Yzerbyt and Kashmina (2005) assessed the relationship between warmth and competence when judging individuals and groups. When individuals were rated a positive correlation between traits was found, with increases in scores of warmth prompting increases in scores of competence and vice versa. However, the opposite was found when groups were rated, where high competence scores were associated with low warmth scores and vice versa. These combinations are proposed to result in stereotypes and can predict prejudice (Fisk, Xu, Cuddy & Glick, 1999).

Work by Oosterhof and Todorov (2008) links the two-dimensional theory of social to perceptions of personality from faces. After asking participants to rate faces on a series of traits Oosterhof and Todorov conducted a principle components analysis, finding two clear factors. Factor one accounted for 63% of the variance, and factor two

18%. All positive judgements, such as attractiveness, had positive loadings and all negative judgements, such as anger, had negative loadings on factor one, suggesting that the factor can be thought of as an evaluation of valence. In addition judgements of trustworthiness were highly correlated to factor one, and judgments of dominance to factor two, however they were not correlated with the other factor, suggesting an orthogonal relationship.

Oosterhof and Todorov (2008) use these findings to suggest that social attribution to faces can be boiled down to two distinct concepts, trustworthiness and dominance, adjectives comparable to Fisk et al.'s warmth and competence dimensions. Again in similar fashion to the concepts proposed by Fisk and colleagues, Todorov, Said, Engell and Oosterhof (2008) propose that dominance is a signal of physical strength, often using cues such as masculinity and maturity, whereas trustworthiness signals valance, whether to approach or avoid and individual, and uses emotional cues. The authors suggest this mechanism is an overgeneralisation of an evolved mechanism, utilising a single judgement of trustworthiness as a proxy for an evaluation of valance. This is supported by the findings that trustworthiness can be accurately judged after just 33ms of exposure to facial stimuli (Todorov, Pakrashi & Oosterof, 2009) and that judgements are pre-conscious (Stewart, Ajina, Getov, Bahrami, Todorov & Rees, 2012).

Oosterhof and Todorov (2009) propose that even neutral facial stimuli convey emotional states, and it is the evaluation of these states which individuals use to attribute personality traits. For example when high and low trustworthy faces portray the same emotional features of happiness, individuals rate the trustworthy face to be happier than the untrustworthy face. This suggests that changes along the dimension of trustworthiness corresponds to subtle changes in facial features relating to valence, or mood. In keeping with this line of thought Todorov and Duchane (2008) demonstrated that ratings of trustworthiness were highly negatively correlated with anger, and positively with happiness, that it highly trustworthy faces were rated as happy, and low trustworthy faces as angry.

Work by Flowe (2012) links this two-dimensional perception of faces to an additional trait, criminality. It was found that upon rating faces of criminals for a range of traits there was a high negative correlation between ratings of trust and criminality and trust and anger, as well as positive correlations between dominance and criminality and dominance and masculinity. Furthering this, when entered into a regression model ratings of trust and dominance significantly predicted criminality scores, suggesting that ratings of criminality can be described using Oosterof and Todorov's two-dimensional concept.

In addition to this Flowe (2012) presented observers with facial images of individuals portraying emotional expressions and asked them to rate the individuals for criminality. In general faces portraying happiness were given a lower criminality rating than faces portraying anger, suggesting once more that valence and trustworthiness can be a proxy for a range of traits including criminality.

In further work Todorov Baron and Oosterhof (2008) have utilised computer models in an attempt to determine which facial features may be used as cues to trustworthiness and dominance. They found manipulating features such as the inner brow ridge of the eyebrows and altering the distance between the mouth and nose had an impact of trustworthy scores. Changes in facial features in this way can be likened to the minor physical abnormalities (MPAs) detailed in Waldrop, Pedersen and Bell's (1968) Waldrop scale. This scale has been utilised in a range of studies finding increased Waldrop scores for individuals with psychiatric disorders, including bipolar spectrum disorder, autistic spectrum disorder and schizotypal personality disorder.

Minor physical abnormalities and psychiatric disorders

One disorder with a considerable degree of research into MPAs is autistic spectrum disorder (e.g. Miles & Hillman, 2000; Rodier, Bryson, & Welch, 1997; Walker, 1977). A meta-analysis of seven studies into MPAs in individuals with autistic spectrum disorder (ASD) were collaborated to demonstrate a pooled effect size of d = 0.84 - suggesting the presence of MPAs in individuals with ASD is a robust finding (Ozgen, Hop, Hox, Beemer, & van Engeland, 2010). In a more recent study Aldridge and colleagues (2011) used 3D photographs of boys with and without ASD in order to demonstrate a distinct phenotype associated with ASD, including the increased breadth of mouth, orbits and upper face alongside reduced height of the philtrum and maxillary (sinus) region.

The presence of increased MPAs has also been noted in schizophrenia, with an early study finding 75% of a sample of schizophrenic males to have anomaly scores above the upper limit of the normal population (Guy, Majorski, Wallace, & Guy, 1983). These results were replicated by Foster-Green, Santz, Galer, Ganzell and Kharabi (1989), who found 31% of a schizophrenic patient sample to have Waldrop scores two standard deviations above the control mean. O'Callaghan and colleagues (1995) found that a majority of these MPAs were found on the head, and that there was a significant correlation between the amount of negative schizophrenia symptoms and number of abnormalities. Utilising a more current approach, geometric morphometric analysis, Hennessy and colleagues (2004) identified a shape variance in individuals with schizophrenia, with patients having more asymmetries than controls.

Finally, the Waldrop scale has been used to try and differentiate between patients with bipolar disorders and controls (Sivkov, Akabaliev, Mantarkov, Ahmed-Popova, & Akabalieva, 2013). Once again, results illustrated a greater prevalence of MPAs in patients, however the results are less clear with affective disorders than with ASD and schizophrenia. Tenyi, Trixler and Csabi (2009) conducted a review of nine studies identified as investigating MPAs in affective disorders. Of the five studies assessing bipolar disorder reviewed, none found a significant increase in MPAs in patients compared to controls, contradicting Sivkov et al.'s findings. A further three studies reviewed concerned unipolar depression, and here only one study reported a significant increase in MPAs in patients.

Links between psychiatric disorders

As previously highlighted, high scores for Neuroticism have been implicated in many neuropsychiatric disorders, and similarly there is consistent evidence for the existence of minor physical anomalies in these disorders. Does the existence of these commonalities between these disorders suggest that there may be a common link between neuropsychiatric disorders? The proposal of a continuum underlying psychiatric disorders was investigated by Crespi, Stead and Elliot (2010), who tested four theories into the association between three common neuropsychiatric disorders, autism, schizophrenia and bipolar disorder. Each of these theories denoted an alternative model for the genomic and etiological relationships between these disorders. The first, the subsumed model, proposes autism is a sub-category of schizophrenia (not distinct to Kanner's original theory of autism), and further assumes an overlap between schizophrenia and bipolar disorder. These three disorders are all grouped away from normality. The second, the separate, model, assumes that schizophrenia and bipolar disorder are associated and on one continuum from normality, and autism is on a separate continuum. In the third, diametric, model, schizophrenia and bipolar disorder are associated on one end of the spectrum, with autism at opposite pole, and normality in the middle. The final, overlapping, model, proposes that all three disorders are associated, overlap to an equal degree, and form the end of a spectrum away from normality. In their paper Crespi and colleagues utilise evidence from a variety of sources, including genomics, neurodevelopmental evidence and phenotypic associations to conclude that the diametric hypothesis is the most appropriate solution.

This provides evidence for Badcock and Crespi's (Badcock & Crespi, 2008; Crespi & Badcock, 2008) diametric hypothesis, which extends the basic diametric model to include additional differences between the groups of disorders and proposes potential origins for them. Autism is proposed to be a consequence of paternally impressed genes, which cause a cognitive style at the mechanistic 'things' thinking end of the spectrum of cognitive style. On the other end of the spectrum is mentalistic 'people' thinking, which Badcock and Crespi propose result in schizophrenia and bipolar (psychotic) disorders and are caused by maternally impressed genes. These can also be thought of as extreme male and female brain disorders, with autism on the masculine end of the spectrum and schizophrenia and bipolar on the opposite, feminine, end.

Other authors have extended upon this theory; suggesting additional disorders including borderline personality disorder (Dinsdale & Crespi, 2013) and eating disorders (Bremser & Gallup, 2012) could be included within the extreme female brain end of the spectrum. These studies suggest that associations of personality traits and MPAs are not the only commonalities between these disorders, but that there may be other underlying continuum of similarities which manifest in symptoms characterised by different cognitive styles.

Recognition of psychiatric disorders from facial images

These studies highlight an apparent difference in the facial phenotype of individuals with and without psychiatric disorders, but are observers sensitive to these differences? and are they able to utilise them to differentiate between individuals with and without disorders? One of the first studies into recognition of personality traits from static facial images started to address this question. Shevlin and colleagues (2003) identified participant's ability to accurately discriminate between individuals scoring high and low for one of Eyesnik's three personality traits, psychoticism. These results were replicated in a more recent study by Holtzman (2011), who utilised a combination of self and peer ratings of the 'dark triad' of personality disorders (psychoticism, narcissism and Machiavellianism) to create composite images from the 10 highest and lowest scorers on each trait. These were then presented to observers in pairs (one high and one low image) who were asked to decide which image they felt best matched the trait description and indicate the confidence they made their decision with. Observers were found to be able to discriminate between the image pairs at a rate significantly above chance levels for all three traits, although post-hoc analysis revealed that accuracy was higher for female images.

These studies have assessed the ability to detect psychiatric personality traits from facial images of individuals scoring high and low on specific traits. Using an adapted methodology Little et al., (2011) collected facial photographs from participants at two time points in the academic year, during the exam period and after a mid-semester break, alongside measures of stress, determined by the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983). These faces were then individually presented to observers, who rated them for stress on a 7-point scale. Participants scored higher for stress during the exam period, a difference which was reflected in ratings from observers. This study highlights not only the ability to detect cues to health from facial images, but also the ability to detect state changes within an individual over time.

Reverting back to a two-alternative-forced-choice methodology, Kramer and Ward (2010) presented observers with composite images created from individuals scoring high and low on the mental component summary of the Short-Form 12-Item Health Survey (SF-12, Ware, Kosinski, & Keller, 1996). However, here, participants were unable to accurately discriminate between images. One possible explanation for this is the lack of specificity in the SF-12, as it gives a general overview of mental health rather than honing in on a specific disorder.

Increased accuracy in specific psychiatric traits is demonstrated by Kleinman and Rule (2013). Over three studies observers were presented with photographs of individuals who had been found to have later committed suicide, alongside control images of individuals who had not. Presented individually, observers were asked to decide if they believed the individual depicted had committed suicide or was still alive. In all studies observers were found to accurately group participants, even when stimuli was cropped to show only the internal facial features. In addition the authors replicated the study using psychotherapists as observers, finding they were able to accurately categorise individuals. However, on average psychotherapists were not significantly more accurate than a community sample of non-clinicians, irrespective of years of experience or level of education. This study highlights the ability for naive observers to detect psychiatric traits with the same degree of accuracy as clinicians, suggesting that the aptitude is within us all.

The current thesis

As demonstrated above there is evidence to suggest that naïve observers are able to detect psychiatric traits when viewing short excerpts of videoed behaviour, even when these clips are reduced to just one second in length. Furthermore, this accuracy is still found when the cues available to observers are reduced even further, presenting just still facial images. However, although limited in length there is still a vast quantity of information available within these clips - body posture, clothing and jewellery to name a few. Additionally, the evidence for this ability relating to specific psychiatric disorders is limited. The first aim of this thesis is to bridge this gap, to determine if naïve observers are able to accurately detect common psychiatric disorders from facial stimuli. This will advance the current literature in three ways; firstly it will serve as a control to determine if accuracy in thin slice studies is a consequence of external cues such as body posture, by using composite still facial image stimuli. Secondly it will extend the existing literature to examine this ability in a wider range of disorders, and finally it will assess these findings in relation to existing theories of psychiatric disorders, to suggest if there is a continuum linking these disorders. In addition the thesis aims to investigate how observers perceive those with symptoms of mental health disorders. Previous research has assessed the ability to detect these symptoms from thin-slice material but has not determined if this perception is correlated with an impression of the individual's personality. This thesis will gather impressions of whether individuals with high or low symptom occurrence have more desirable personality traits, with the aim of shedding light on how individuals with mental health symptoms are perceived and the effect this may have on social interactions.

This thesis contains seven experimental chapters; in chapter 2 I assess the ability to detect depressive symptoms from static facial images, as well as exploring the possibility that perceptions of depressive symptoms may affect personality trait attributions. In chapter 3 I follow this up by reducing the stimuli to determine where in the face this information is portrayed. Chapters 4, 5 and 6 extend these findings, determining if observers are able to detect schizotypal personality disorder, borderline personality disorder and autistic spectrum disorder respectively. Once more I also explore the potential effects of these ratings on personality attributions, paying particular attention to the social implications of these inferences and the role they play in symptom recognition. In chapters 7 and 8 I develop these findings to test existing theories of the basis of psychiatric disorders. In chapter 7 I assess Baron-Cohen's masculinity hypothesis of autism, determining whether individuals scoring highly on the Autistic Quotient are perceived to be more masculine in facial appearance. I additionally review the current evidence for long-term implications of pre-natal testosterone levels on facial development. In chapter 8 I extend these findings to reflect Badcock and Crespi's diametric hypothesis, assessing if females with 'psychotic spectrum' disorders such as bipolar disorder are perceived to be more feminine in facial appearance - the polar opposite of Baron-Cohen's masculinity hypothesis. These chapters not only investigate support for existing theories of the underpinnings of neuropsychiatric disorders, but also assess the possibility of commonality between disorders, and how this could affect the accurate recognition of such disorders from facial images.

Chapter two: Facial cues to depressive symptoms and their associated personality attributions[1]

Abstract

Depression is a common mental health disorder, with 12% of the UK population diagnosed at any one time. We assessed whether there are cues to depressive symptoms within the static, non-expressive face, and if other socially-relevant impressions might be made by these cues. Composite "average" face images were created from students scoring high and low on self-report measures of depressive symptoms, capturing potential correlations between facial appearance and symptoms of depression. These were then used in a warping procedure, creating two versions of individual faces, one warped towards the high symptom composite, and the other towards the low. In Experiment 1, we first found observers were able to identify images representing high and low symptom occurrence at levels significantly greater than chance. Secondly, we collected observer impressions of the two versions of each face. The faces reflecting high levels of depressive symptoms were picked as less socially-desirable over a broad range of personality trait estimates compared to low symptom images. In Experiment 2, we replicated the key finding that the static face contains cues to levels of depression symptoms, using composites created from a new database of student photos and depression inventory scores. Social implications of these perceptions are discussed.

---

[1] This chapter appears in print:
Scott, N. J., Kramer, R. S. S., Jones, A. L., & Ward, R. (2013). Facial cues to depressive symptoms and their associated personality attributions. *Psychiatry Research, 208, 47-53.*

2.1.Introduction

Depression is a disabling mental disorder which affects 12% of adult UK population at any one time (Singleton, Bumpstead, O'Brien, Lee, & Meltzer, 2001). The effect of this widespread diagnosis is far-reaching; by 2030, depression is proposed to be the second largest cause of worldwide disability, rising to the leading cause in high-income countries (Mathers and Loncar, 2006). At an individual level, those with depression are more likely to have fewer qualifications, a lower socioeconomic status (Singleton et al., 2001), and form a group that have a higher suicide incidence rate than the general population (Simon and Vonkorff, 1998). Depression also has a negative impact on the economy, with patients requiring more doctors, community care and medication than others (Singleton et al., 2001). A deeper understanding of the ability to discriminate between individuals with and without depressive symptoms, alongside how this affects perceptions of social desirability may aid knowledge of the social implications of perceived depression.

There are two main types of cues that could be utilised to help discriminate between depressed and non-depressed individuals. The first considered here is that of live cues, where participants are able to use cues from body motions. In a study of "thin slices" of non-verbal behaviours and appearance, Waxer (1976) showed two-minute silent video segments of depressed and non-depressed patients to undergraduate and clinical graduate students. The patient's Minnesota Multiphasic Personality Inventory depression subscale scores and participant's depression ratings showed a correlation of 0.60, suggesting an ability to recognise depressive symptoms from non-verbal cues. These live cues included information from physical movements such as body motion, facial expressions, posture and proximity to others which could help to identify individuals with depression. Use of these live cues has been explored in other studies;

for example, Michalak et al. (2009) found that body motion in minimal point-light displays was a reliable cue to depression, with depressed participants walking more slowly and with less amplitude in limb movements. Geerts and Brune (2011) discuss findings of other live cues of depression, such as the ability to predict depression course from the impaired social behaviour of depression-prone patients.

However, recent research indicates an unexpected richness from a second type of cue - static cues. An example is the static, non-expressive face. Many of the Big Five personality traits, as measured by self-report, can be accurately estimated from "passport"-type photographs of people looking straight-ahead and with a neutral expression, both using individual photos (Penton-Voak, Pound, Little, & Perrett, 2006) and composite images comparing the highest and lowest trait levels (Kramer and Ward, 2010; Little and Perrett, 2007). Composite images are computer-generated averaged faces of a number of individuals. These composite images reduce individual variation and preserve the common facial features of those within the group. Besides Big Five personality traits, the face can also reliably cue socially-relevant traits such as sociosexual orientation (Boothroyd, Jones, Burt, DeBruine, & Perrett, 2008) and trustworthiness (Stirrat & Perrett, 2010).

Limited research has been conducted into the ability to recognise clinically-related conditions from facial images. Shevlin, Walker, Davies, Banyard and Lewis (2003) demonstrated that participants were able to accurately judge levels of psychoticism from facial photographs. However as jewellery, clothing and hairstyles of the stimuli were not kept consistent it cannot be determined whether the accuracy reported is a product of facial cues, or aided by additional cues such as clothing type. Research by Holtzman (2011) has helped to overcome these criticisms by using images in which individuals have removed all make-up and jewellery, and tied their hair away from their face. Even after removing these confounds participants were able to accurately identify the ‘dark triad’ of personality disorders (machiavellians, narcissists and psychopaths) from composite facial images.

Results like those above raise the question of whether reliable cues for symptoms of depression (and other mental health conditions) may be present in the non-expressive face. The possibility of a link between facial appearance and symptoms of depression is strengthened by two findings. First, high trait Neuroticism is a risk factor for depression; and second, trait Neuroticism is one of the factors reliably signaled on the face. We discuss these findings in turn.

Personality traits are heritable constructs that serve as risk factors for mental health disorders (Kruger, Caspi, & Moffitt, 2000; Sen et al., 2003). Neuroticism has been identified as a personality trait with special importance for depression, as it may increase the likelihood of an individual experiencing distress (Bagby et al., 1996), which can in turn manifest itself as anger, anxiety or depression (Bagby et al., 1997). Indeed, individuals with depression are consistently found to have higher than average levels of Neuroticism (Bagby et al., 1997; Trull & Sher, 1994; Widiger & Trull, 1992). Although there is controversy surrounding the direction of the neuroticism-depression link, the general consensus suggests that Neuroticism forms a diathesis for depression (Krueger, Caspi, & Moffitt, 2000; Trull & Sher, 1994). The ability to use Neuroticism levels to predict later depression provides some supporting evidence for this theory (Roberts & Kendler, 1999).

Neuroticism also seems to be one of the traits which is most reliably signalled from the neutral face. Kramer and Ward (2010) found that composite images of women with high and low levels of trait Neuroticism could be identified at levels significantly greater than chance, even from internal facial features alone. In fact, Kramer and Ward

(2010) found that removing the outer facial features (top of the head, jaw-line, etc), leaving only the eyes, nose, and mouth areas within the images, did not impair identification of Neuroticism. More recently, Jones, Kramer and Ward (2012) created high and low Neuroticism composite images on the basis of 3D facial scans, to remove potential postural or alignment cues from subsequent comparisons. Again, identification of Neuroticism composites was robust even without these cues.

We sought to determine whether symptoms of depression might be similarly cued in the static, non-expressive face in two experiments. In Experiment 1, we tested the possibility that, within a non-clinical sample, people who score high for symptoms of depression have a distinct facial appearance from those who score low for depressive symptoms, and if so, whether such differences in appearance can be identified by untrained observers. We also explored how appearance impacted observer impressions of "social desirability", that is, whether the person depicted in the photo was perceived to be a desirable social partner. In particular, we examined whether the facial appearance correlated with depressive symptoms might leave negative impressions on observers. These observer impressions might be important even if they are not accurate, as they could mean that people at risk of depression receive unfavourable social reactions, even in the absence of overt depressive behaviour. This unfavourable reaction could then increase social isolation and negative social consequences associated with depression. In Experiment 2, we created a new database of photographs and depression inventory scores, to replicate the identification task used in Experiment 1, with a new set of composite images.

Experiment 1

2.2 Method

This experiment consisted of two phases, the first of which ("stimulus creation") involved the collection of photographs and a measure of depressive symptoms to create our face stimuli. In the second phase, these stimuli were used to investigate the ability of observers to discriminate between faces representing high and low depression symptom occurrence (Experiment 1a), and to measure observer impressions from the high and low images of important social traits (Experiment 1b). All stages of the study were approved by Bangor University's departmental ethics committee.

2.2.1. Stimulus creation

Our aim in this phase was to create stimuli which reflected the actual differences, if any, in the faces of people who report high and low levels of depressive symptoms. To do so, we created composite ("averaged") images from the faces of people who reported they had high levels of depressive symptoms, and another set of composites from the faces of people reporting low levels of symptoms. These composites were then used as anchor points to let us create new images, by warping individual faces towards the high and low depression appearances.

2.2.2. Participants: Two hundred and twenty five Bangor University students (130 females, age $M = 21.45$, $SD = 5.04$) took part in photo shoot A (See appendix B for details) and paid £5 for their participation.

2.2.3. Measures: Depressive symptoms were assessed using questions based on the revised Beck's depression inventory ([BDI], Beck, Steer, & Brown, 1996), in which depression symptom severity was indicated on a 4-point Likert scale ranging from 0 (disagree) to 3 (very much agree). Summed scores on the BDI can give an indication of depressive symptom severity, with scores between 0-13 indicating minimal depression, 14-19 mild, 20-28 moderate and 29-63 severe depression (Beck et al., 1996). Question nine, asking participants about suicidal thoughts, was removed in accordance with the Bangor University's departmental ethics committee, resulting in a 20 item questionnaire with a possible score range of 0-60. This adaption did not appear to impact upon the reliability of the questionnaire (Cronbach's $\alpha = 0.91$). Male scores ranged from 0-49, $M = 14.94$, $SD = 12.36$, female scores ranged from 0-43, $M = 16.41$, $SD = 10.53$. These scores were later used to separate individuals reporting the highest and lowest levels of depression symptoms; purely for simplicity of exposition, these will be referred to as depressed and non-depressed groups, although to be clear, these were both non-clinical samples.

2.2.4. Procedure: Facial photographs were taken with a professional camera from a distance of two metres whilst camera height, zoom and flash were kept consistent. Participants were asked to sit down in front of a clear, light coloured background and remove make-up and jewellery, tie their hair back and adopt a neutral facial expression with eyes facing the camera. Individuals whose self-reported ethnicity was not White British, those aged over 30, and males with facial hair were removed in order to maintain composite quality, resulting in a sample of 106 females and 48 males.

The fifteen highest (female score $M = 34.73$, $SD = 5.60$, male $M = 27.73$, $SD = 9.07$) and lowest scorers (female score $M = 2.53$, $SD = 1.46$, male $M = 3.13$, $SD = 2.23$) on the depression survey were then selected for each sex, and used to produce high and low depression composites for each sex (Figure 2.1) using JPsychoMorph software (Tiddeman, Burt, & Perrett, 2001).

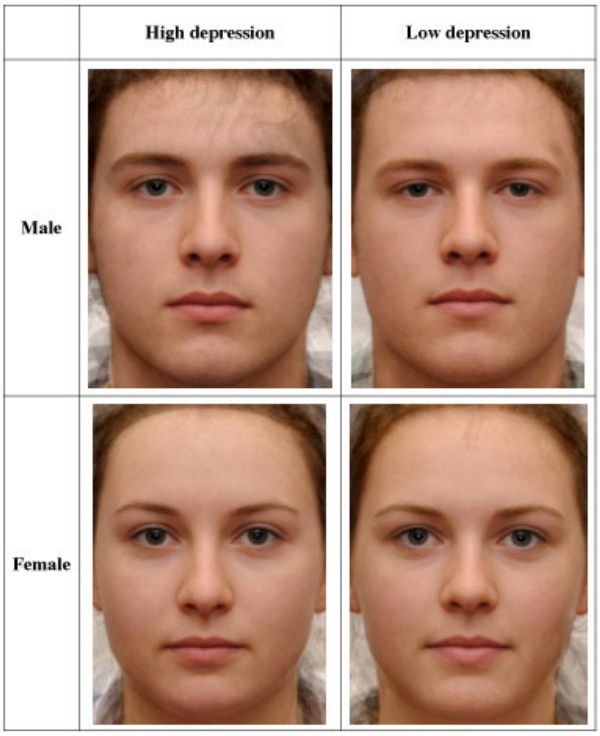


Figure 2.1. Composite images created from the photographs of the highest and lowest 15 male and female depression scorers.

The composite faces shown in Figure 2.1 capture statistical regularities in the facial appearance of people who reported high and low levels of depressive symptoms. These composites were not shown to observers, but were used in the next part of the stimulus creation process, as anchor points to create sexually-specific gradients of shape and texture changes related to high and low depression. Twenty male and twenty female individuals with a depression score close to the group average, and who had given consent to have their photograph used individually, were picked at random from the photograph dataset. The high and low composite images were used as anchor points on two ends of a continuum and each individual's image was warped 50% towards the high and also 50% towards the low depression composites (Rowland & Perrett, 1995). This resulted in two images for each individual, one transformed to reflect the statistics of face shape and texture in the high depressed composite, the other the low (see Figure

2.2 for examples). The 40 stimulus pairs created were used in the following discrimination experiments.

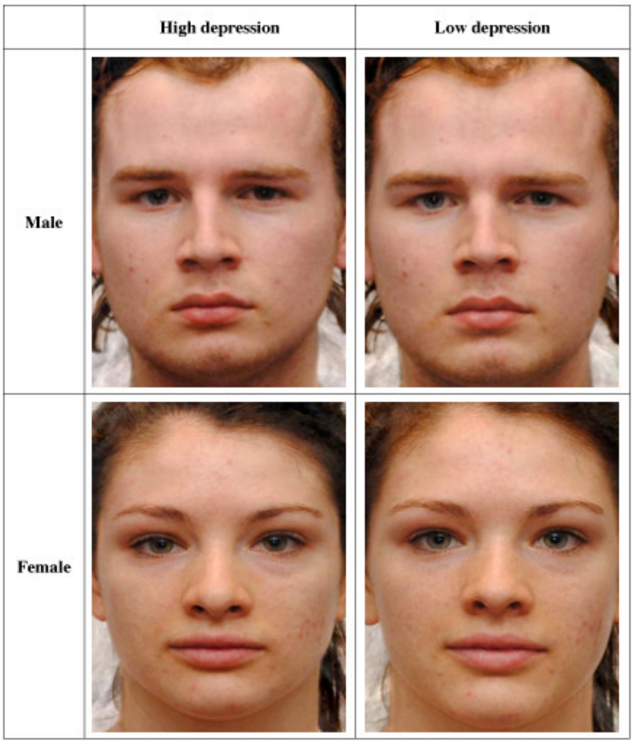


Figure 2.2 Warped image pair examples. Images warped 50% towards the high and low depression composite images illustrated in figure 2.1.

2.3. Experiment 1a: Can depressive symptoms be identified from the static face?

In this experiment, we assessed observers' ability to accurately discriminate between the 40 facial image pairs created in Phase One. Within each pair, a facial photo was transformed to produce one image reflecting the statistics of face shape and texture relating to high levels of depressive symptoms; the other, low levels.

2.3.1 Method

2.3.1.1. Observers: Sixty-four undergraduate Bangor University students aged 18-41 (46 females, age *M* = 19.90, *SD* = 4.03) participated in the experiment for course credit.

2.3.1.2. Procedure: Observers were presented with each warped image pair (37.6 x 50.0mm on screen) side by side (position counterbalanced) on a computer screen below a discriminative statement. Observers were asked to select which image had ‘more symptoms of depression’ for all image pairs. Responses were not speeded, and viewing distance was not fixed. After selecting an image to match the discrimination statement using a mouse click, the next trial was presented. The 40 trials were presented in a randomised order for each observer.

2.3.1.3. Statistical analysis : A 2 (sex of observer) × 2 (sex of stimuli) mixed ANOVA was conducted to determine if the sex of observer, sex of stimuli or an interaction between the two had significant impact on discrimination accuracy. Following this a one-sample t-test was used to compare observer accuracy to chance performance. Chance performance in this two-alternative, forced-choice discrimination task was 50%.

2.3.2 Results

Correct responses were counted when observers selected the high depression image as having more symptoms of depression. A 2 × 2 mixed ANOVA showed no significant effect of stimulus sex, $F(1, 62) = 0.74$, $p = 0.40$, observer sex, $F(1, 62) = 0.18$, $p = 0.67$, or interaction effects, $F(1, 62) = 0.30$, $p = 0.86$, suggesting that there was no difference in accuracy between male and female observers, or between male or female stimuli. For these reasons results are presented aggregated over stimulus and observer sex. Observers were able to accurately discriminate high and low depression images 61%, 95% CI [60-63%], of the time, a level significantly greater than chance, $t(63) = 9.81$, $p < 0.000001$, $d = 1.19$.

2.4. Experiment 1b: Perceptions of social traits in faces high and low for depressive symptoms

Experiment 1a showed that observers were able to discriminate between images reflecting the faces of people with high and low levels of depressive symptoms more accurately than expected by chance. Given that untrained observers were sensitive to the correlation between facial appearance and depressive symptoms, it is interesting to ask what other perceptions these faces might afford. In Experiment 1b we explored the impressions of socially-relevant traits attributed to faces representing high and low depression. That is, what kinds of social impressions would observers receive from a face reflecting the statistical properties of high or low levels of depressive symptoms? For example, would observers have the impression that depressed faces were less emotionally stable or agreeable? These impressions made by the face are relevant as they may bias the social responses of observers. We sought to determine if pre-assessed correlations between depression and personality traits would be reflected in perceptions by observers, such as higher scores in Neuroticism and lower in Extraversion and Agreeableness. We assessed perceptions of Big Five personality traits, as a summary description of socially-relevant impressions. Attractiveness and perceived gender were also included as important socially-relevant variables that are driven by appearance.

2.4.1 Method

2.4.1.1. Observers: A new sample of 48 Bangor University undergraduate students aged 18-33 (37 females, age M = 19.78, SD = 2.39) took part in the experiment for course credit.

2.4.1.2. Procedure: The 40 pairs of high and low depression images were presented in a similar way to Experiment 1. In this case however, each pair was presented with seven different discrimination statements, for a total of 280 trials. The seven discrimination statements consisted of one statement for each of the Big Five personality factors, one for attractiveness, and one for perceived femininity. For most cases, apart from gender and Neuroticism, observers were asked to select the face which had the more socially desirable appearance: more Extraverted (“Who looks more outgoing and energetic?"), Agreeable ("more friendly and compassionate"), Conscientious ("Who looks more efficient and organised?"), Open ("Who looks more open to new experiences?"), and attractive ("Who is more attractive?"). For the Neuroticism discrimination, observers were asked "Who looks more sensitive and nervous?" and for the gender discrimination observers were asked to select the face which was more feminine. This final discrimination statement was chosen to explore the perceived gender of the stimuli, assuming masculinity and femininity are at opposite ends of a continuous spectrum.

To reduce chances of fatigue the experiment was separated into two blocks, with a short break offered in-between. One block contained the female images and the other the males, presented in a counterbalanced order across observers. The blocks were split in this way as the male and female stimuli were created separately, based on different sets of anchor points. Within a block, face pairs and discrimination statements appeared in randomised order for each observer. Between blocks, a break was given as needed.

2.4.1.3. Statistical analysis: A 7(Discrimination statement) × 2(Sex of stimuli) × 2(Sex of observer) factorial MANOVA (robust to violations of sphericity) was used to determine the effect of observer and stimuli sex on observations. Following this one-sample t-tests were used to compare observers accuracy to 50% chance level. Due to the use of multiple t-tests a Holm's sequential Bonferroni correction (Holm, 1979) was applied, resulting in a stricter alpha level and reducing the chance of type-one errors. Additionally correlation analysis was used to determine the possible influence of perceived attractiveness; again a Holm's sequential Bonferroni correction was used to compensate for multiple tests.

2.4.2. Results and discussion

For easy comparison, scores for Neuroticism were reversed to represent Emotional Stability, so that all personality traits now increase in desirability as scores increase.

A 7 × 2 × 2 factorial MANOVA showed only the expected main effect of trait, $F(6, 41) = 4.28$, $p = 0.002$, but no significant effect of stimulus sex, $F(1, 46) = 1.67$, $p = 0.69$, observer sex, $F(1, 46) = 0.18$, $p = 0.67$, or interaction between the two, $F(1, 46) = 0.28$, $p = 0.60$, or any other interaction, all F's$(6, 41) < 1.80$, $p > 0.05$. We therefore carried out subsequent analyses separately for each trait, aggregating over stimulus sex and observer sex.

As illustrated in Figure 2.3, when presented with high and low symptom versions of the same face, observers attributed more positive personality traits to the facial images of low depression, at a rate significantly higher than expected by chance. Compared to high depression images, low depression images were selected as more Agreeable (58%, 95% CI [55-60%]), $t(47) = 5.58$, $p = 0.00000069$, $d = .81$, Conscientious (53%, 95% CI [50-56%]), $t(47) = 2.41$, $p = 0.0199$, $d = .35$, Emotionally stable (58%, 95% CI [55-61%]), $t(47) = 5.07$, $p = 0.0000034$, $d = .73$, Extraverted (58%, 95% CI [54 - 61%]), $t(47) = 4.26$, $p = 0.00039$, $d = 61$, and Open (60%, 95% CI [57 - 63%]), $t(47) = 7.26$, $p = 0.000000023$, $d = 1.05$, corrected for multiple comparisons.

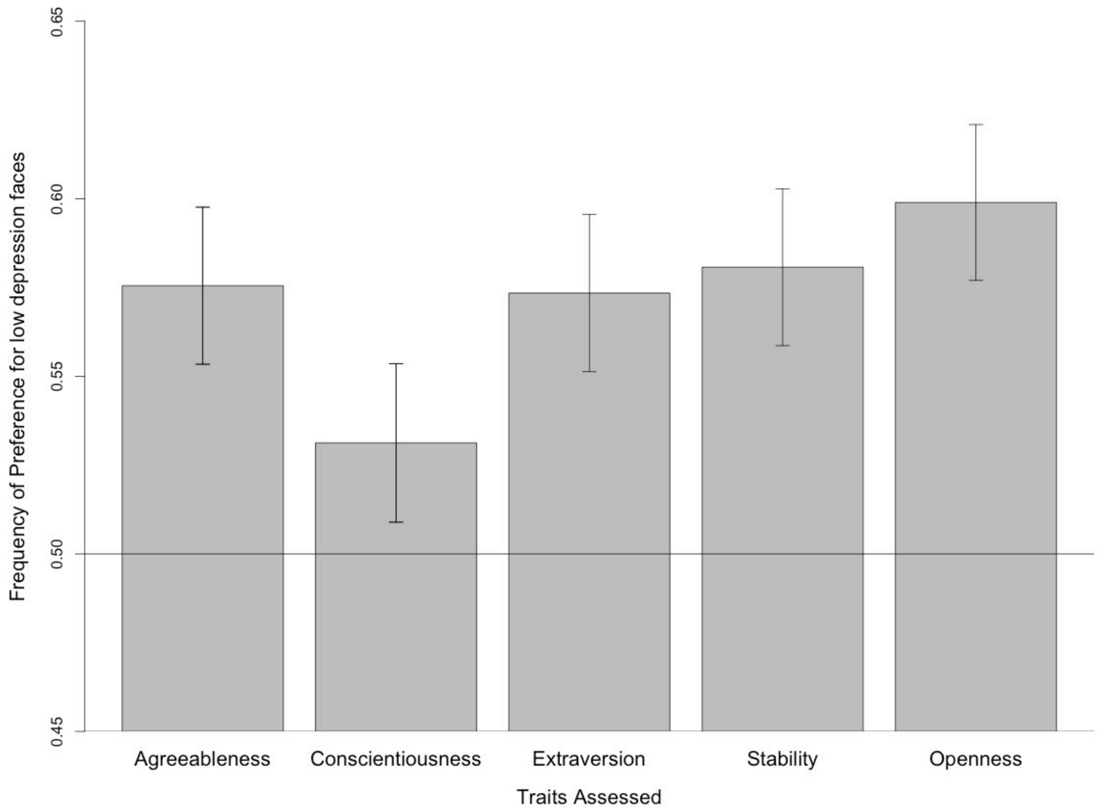


Figure 2.3. Frequency of preference for low compared to high depression faces. Error bars represent 95% confidence interval of the mean for each trait.

When compared to high depression images, observers also perceived the low depression images to be more attractive (55%, 95% CI [52 - 57%]), and feminine (55%, 95% CI [52 - 58%]), of the time, again a rate significantly greater than expected by chance, $t(47) = 3.30$, $p = 0.0054$, $d = .48$, and $t(47) = 3.24$, $p = 0.0044$, $d = .56$, (corrected for multiple comparisons) respectively. Attractive individuals are often perceived to be high in desirable traits, a concept known as the halo effect (Dion, Berscheid, & Walster, 1972). As low symptom occurrence faces were perceived to be more attractive, as well as higher in desirable personality traits, we conducted a series of correlation analyses to determine if the halo effect was the driving force behind our findings. Although correlations were generally positive, only perceptions of Openness and attractiveness were found to be significantly correlated, $r = 0.62$, $p = 0.00002$, all other correlations were non-significant , all r's < .34, all p's > 0.08, corrected for multiple comparisons[2]. Despite the association between perceived levels of Openness and attractiveness, the absence of significant correlations between the other personality traits and attractiveness suggests that an attractiveness halo is unlikely to be the driving force behind the wide-spread attribution of more negative traits to the high depression images.

Experiment 2

2.5. – Identifying levels of depressive symptoms from the static face in a new sample

In Experiment 1 we demonstrated that observers were able to identify at above-chance levels facial images representing high and low symptoms of depression. We thought it was important to replicate and generalise this finding. We sought to replicate these results by creating a new database of faces images, and generating new high and low composite and warped images. We also wanted to be certain that our results were not sensitive to the depression inventory used, and so this time levels of depressive symptoms were measured using the Inventory of Depressive Symptoms (IDS, Rush, Gullion, Basco, Jarrett, & Trivedi, 1996).

2.5.1. Method

This experiment was run using the same procedures as Experiment 1a, with the differences noted below.

2.5.2. Stimulus creation: Stimulus creation proceeded exactly as in Experiment 1, with the following modifications. First, a new set of two hundred and twenty-one

---

[2] Please note the small sample size in this study results in a loss of power, possibly affecting the ability to determine a true result.

Bangor University students (130 females, age *M* = 21.65, *SD* = 5.09) took part in photo shoot B (See appendix B for details) and paid £6 for their participation. Second, the Inventory of Depressive Symptoms (IDS, Rush et al., 1996) was used as the measure of depressive symptoms. The IDS is a 30 item, self-rated, questionnaire where participants are asked to describe symptom severity on a 4-point scale, ranging from 0 suggesting no symptom presence, to 3 suggesting high symptom presence. Possible scores range from 0-84, with scores 0-25 suggesting no depression, 26-38 mild depression, 39- 48 moderate depression and 49-84 severe depression. Female scores ranged from 1-62 (*M* = 17.48, *SD* = 11.56), male scores ranged from 1-43 (*M* = 15.41, *SD* = 9.03). The images of the fifteen highest (female score *M* = 30.40, *SD* = 8.40 , male *M* = 25.40, *SD* = 7.69) and fifteen lowest (female score *M* = 5.40, *SD* = 1.74, male *M* = 6.40, *SD* = 2.36) scorers on the IDS were selected and then used to create composite stimuli for each sex. Following this 20 male and 20 female images were selected to create the warped images, using the same procedure as in Experiment 1.

2.5.2.1 Observers: Forty-eight (36 females) undergraduate Bangor University students took part in the study for course credits. Observers' ages ranged from 18-36 (*M* = 20.54, *SD* = 3.05).

2.5.2.3. Procedure: The experimental procedure was identical to Experiment 1a.

2.5.2.4. Results and discussion

Answers were coded as correct when observers selected the image representing high depressive symptoms. A 2 (Sex of stimuli) x 2 (Sex of observer) mixed ANOVA showed no significant effect of observer sex, $F(1,46) = 1.05$, $p = 0.31$, stimulus sex, $F(1,46) = 0.45$, $p = 0.50$, or interaction, $F(1,46) = 0.04$, $p = 0.84$. For this reason means were aggregated over stimulus and observer sex. Observers were accurate in their discriminations 58%, 95% CI [54-62%], of the time, a rate significantly more accurate than expected by chance *t*(47) = 3.86, *p* = 0.0003, *d* = .56. An independent samples t-test found no difference in accuracy between Experiment 1a and 2, *t*(110) = 1.68, *p* = 0.10, *d* = 0.32. Observers were therefore consistently able to identify facial images of individuals scoring high and low on measures of depression, significantly more often than expected by chance. These findings replicated over two independent stimulus samples, different depression measures, and different observer samples.

2.6 General discussion

We created images which captured statistical regularities in the facial appearance of students who did and did not report high levels of depressive symptoms -- less formally, depressed and non-depressed faces. We found in two separate studies that untrained observers were able to discriminate between these images more accurately than expected by chance. These results demonstrate two important conclusions about facial appearance and depression. First, there are differences in the facial appearance of people based on their reported levels of depressive symptoms. Second, observers are sensitive to these differences: not only were observers able to discriminate between the high and low depression images more accurately than expected by chance, but they also made more negative social attributions towards the high depression images.

One range of possibilities is that there are biological factors, for example, genetic or hormonal, which influences both susceptibility to depression and facial appearance. Additional possibilities are related to a "self-fulfilling prophecy", in which people with a particular appearance learn through interaction with others to behave in a particular way. Here we briefly consider both possibilities.

The biological basis of depression is under active study, yet heritability estimates of Neuroticism are substantial, in the range of 40% (Lake, Eaves, Maes, Heath, & Martin, 2000) to 50% (Floderus-Myrhed, Pedersen, & Rasmuson, 1980). Interestingly depression has also been shown to have a genetic aspect, with an estimated heritability of 31-42%, and additional findings that first-degree relatives of patients are more likely to have recurrent depression (Sullivan, Neale, & Kendler, 2000). It is possible that this is a reflection of the genetic transmission of Neuroticism that puts individuals at increased risk of depression development (Bagby et al., 1997). The type of depression discussed here would be seen as trait depression, where an individual has a stable trait (such as Neuroticism) which increases the likelihood of experiencing severe depression. However, depression can also occur as a state – where symptoms fluctuate over time (Teasdale, 1998). It is often hard to tease apart these disorders, however recent research conducted by Endler, Macrodimitris and Kocovski (2000) suggests that measures such as the Beck's Depression Inventory (Beck et al., 1996) are accurate in measuring trait levels of depression. This means that the stimuli developed in this study are more likely to represent those who experience trait, as opposed to state, depression.

It is also clear that the face is a sensitive marker for many kinds of congenital and acquired disorders. Facial appearance is a diagnostic criterion in many such disorders (e.g., Downs and foetal alcohol syndromes). Williams syndrome is an example of a genetic disorder which presents a behavioural pattern of hyper-sociality alongside distinct morphological facial features (an 'elfin' shaped face). Although such results do not speak to our specific case of depression and the face, they do demonstrate that behavioural traits and facial morphology can be correlated in a highly specific manner. In short, a genetic or hormonal link between depressive symptoms and facial appearance seems like a plausible story, but there is little direct evidence for it at this point.

Our results bear more directly on the possibility of a cultural or acquired link between appearance and depression. Personality traits are important factors in determining friendship and social inclusion. Friends of highly extraverted individuals rated their friendship as closer, while friends of highly agreeable individuals reported less irritation and friendships with highly open individuals had fewer reported conflicts (Berry, Willingham, & Thayer, 2000). Furthermore, individuals scoring highly on Agreeableness had higher numbers of new acquaintances wanting to form friendships with them (Selfhout et al., 2010). This literature suggests that individuals high in these personality traits (Agreeableness, Extraversion and Openness) are seen as more desirable friends, traits that which observers perceived to be less prevalent in individuals with high symptom occurrence. Given this evidence, it is highly relevant that we find the social response to the faces of people with high levels of depressive symptoms to be generally negative: for example observers perceived images of high depressive symptoms to be less Agreeable, less Conscientious, less Extraverted, less Open to experience, and less Emotionally Stable. That is, the high depression faces seem to be involuntarily broadcasting a message of negative personality traits associated with social undesirability, even in the absence of any overt behaviour. At the very least, this would mean that people with a prototypical high-depression face would face barriers and challenges for social inclusion that others would not. A failure to meet these social challenges could heighten any risk of depression, by making social exclusion more likely. This means that an individual's depressive symptoms may cause them to feel like they are being excluded socially, whilst the facial cues picked up by others casts a negative social light on them. Indeed findings from Segrin (1992) that mild symptoms of depression can induce rejection from others further support this argument.

Our results demonstrate that facial appearance can hold cues for symptoms of depression. Additionally we have shown that observers perceive the facial appearance associated with depressive symptoms to be less socially desirable reflecting previously identified associations with big five personality traits and depressive symptoms. There are limitations to our study which are important to highlight. The photograph databases are collected from a non-clinical student sample, so it will be important to investigate the association between facial appearance and high-levels of depression symptoms in a clinical population. This would help to determine the extent to which facial appearance is a useful cue for identification of depression. Secondly, although discriminations between high and low depressive symptom images are more accurate than expected by chance, performance is far from perfect. However, there are evidently cues within the face correlated to levels of depressive symptoms which further knowledge about the correlation between appearance and mental health, and what the visible cues may be that underlie this correlation.

In practical terms, given that observers formed negative impressions from the facial appearance statistically associated with depression, it may be that facial appearance variables are associated with social outcomes. For example, a face which approaches the high-depression average may be seen as having fewer socially desirable personality traits, which in turn could negatively effect social interactions.

The next step

In this chapter I showed that participants were able to accurately detect depressive symptoms from static facial images. In the next chapter I investigate whether it would be possible to locate which facial region is utilised to aid this ability.

Chapter three: Depressive symptoms expressed within the eyes and internal facial features: Thin slices cueing depression

Abstract

Previous research has demonstrated that untrained observers can accurately detect depressive symptoms solely from static facial images (Scott, Kramer, Jones, & Ward, 2013). However, the cues used to make these discriminations are currently unknown. To narrow this down, in this chapter it was determined if symptoms could be accurately identified even when cues were limited to internal facial features and just the area of the eyes was investigated. Stimuli from Scott et al. (2013) were cropped to create two stimuli sets; internal features (eyes, nose, mouth) and just the eye region, which were then presented in pairs to observers who were asked to choose which image had more or fewer symptoms of depression, as well as personality attributions. Observers were found to accurately discriminate between the high and low depressive symptom images in 58% of trials with internal features and 56% when just eyes were shown. These findings suggest that the eye region, including the eye brows, hold important cues in the discrimination of depressive symptoms in the face. This may be a result of cues from structural features, from expressive features such as eyebrow position, or a mixture.

3.1 Introduction

The static, neutral face has been demonstrated to be a rich source of social information, with studies showing accurate recognition of traits ranging from personality (Kramer & Ward, 2010; Little & Perrett, 2007; Penton-Voak, Pound, Little, & Perrett, 2006) to trustworthiness (Stirrat & Perrett, 2010), suicidality (Kleiman & Rule, 2013), and depression (Scott, Kramer, Jones, & Ward, 2013). For some traits, the specific parts of the face used by observers to make their judgments have been identified. Santos and Young (2011) occluded different regions of the face and measured observer agreement for different social traits. They found, for example, consensus for approachability and trustworthiness from just the internal facial features (eyes, nose, mouth, but not jaw or facial outline).

An important distinction can be made between studies which define accuracy by whether observer judgments agree (consensus, such as measured by Santos & Young, 2011), and whether those judgments are accurate. The accuracy of social trait judgments can be assessed by measuring whether observer perceptions from facial images are in agreement with the actual social traits of those pictured. Again, occlusion of facial features can be used to identify the areas used to make accurate personality discriminations. Kramer and Ward (2010) measured the accuracy of personality judgments from the face using full-face images, and from images cropped around the internal features of the eyes, nose and mouth. Kramer and Ward found that observers were able to accurately identify four (Agreeableness, Extraversion, Openness and Neuroticism) of the big five personality traits from internal features only. In one case (extraversion), accuracy was reduced relative to the full faces, but in another case (openness), accuracy improved for the internal relative to the full faces. When investigating the ability to accurately determine if an individual was to go on to commit suicide Kleiman and Rule (2013) demonstrated that observers were still accurate in their categorisation when faces were cropped to only show the internal facial features. These studies suggest that the internal facial features hold important cues to social traits, however additional research suggests that observers still maintain accuracy when this information is further reduced.

Tskhay, Feriozzo and Rule (2013) measured observer accuracy for determining female sexual orientation. In this study, observers were asked to judge a female face as ‘straight’ or ‘lesbian’, and these judgments were compared to the reported sexual orientation of those photographed. Images presented included full faces, internal features only, and images encompassing just the eye region. Observer’s accuracy was above chance levels in all conditions, results replicating earlier work (Rule, Ambady & Hallett, 2009), although as visual information was reduced the accuracy level also decreased. Rule, Garrett and Ambady (2010) took a similar methodology when assessing the ability to categorise Mormons and non-Mormons from facial stimuli. Here the authors demonstrated an above chance accuracy even when observers were presented with facial stimuli which had the hair cropped from the image, and the eyes and mouth blanked out. Interestingly when presented with the internal features (eyes, nose and mouth) individually, observers were unable to accurately categorise the features.

These studies highlight the different facial cues utilised across the perception of social traits. This chapter seeks to further the results of chapter 2, to assess the ability to discriminate between high and low depressive symptoms when stimuli is reduced in a similar methodology to Tskhay et al. (2013). It was hypothesised that internal facial features alone may allow above-chance discrimination of high and low levels of depressive symptoms, as Kramer and Ward (2010) demonstrated above-chance accuracy for trait Neuroticism using this type of stimuli. Depression is associated with increased Neuroticism scores (Widiger & Trull, 1992). It has been suggested that trait Neuroticism could form a diathesis for depression (Trull & Sher, 1994) as an individual with higher levels of Neuroticism is more likely to experience distress (Bagby, 1996), which can manifest itself into depression (Bagby, 1997). Additionally those with high Neuroticism scores are more likely to have poor coping and defense skills and thus are more affected by life situations (Ormel & Wohlfarth, 1991). In addition to presenting internal facial features observers were asked to make discriminative judgments based on the eye and brow region. It was hypothesised that if observers were accurate in judgments based on internal features this may extend to the eye region as it has previously been implicated as a cue in the detection of Neuroticism, with individuals with larger eyes being rated as higher in Neuroticism (Paunonen, Ewan, Lefave & Goldberg, 1999). The use of the eye region as a cue to social traits is not suprising when it is considered that on viewing facial stimuli observers have been shown to spend an average of 43.5% of their time focusing on the eye region (Janik, Wellens, Goldberg & Dell'osso, 978)

To investigate whether the accuracy in determining depressive symptoms from facial images demonstrated by Scott et al. (2013, chapter 2) could be explained by cues from specific parts of the face we created two new versions of Scott et al.'s stimuli, first by blocking out external features including the jaw line leaving a "T" shaped stimuli encompassing the internal facial features of the eyes, nose and mouth. In the second stimulus set, similar to Tskhay et al. (2013), observers were presented with an even more impoverished stimulus, removing all features except the eye region. These sets of stimuli were then presented to observers in a two-forced choice experiment to determine if individuals are still able to accurately differentiate between facial images of high and low depressive symptoms when fewer cues are available. In addition, Scott et al. found observers to attribute negative personality traits to individuals with high levels of depressive symptoms. In a second study we sought to determine if these personality attributions were upheld with the reduced stimuli.

3.2. Method

3.2.1. Stimuli

Facial images were developed from the set created by Scott et al. (2013). Scott et al. created composite images from students reporting high and low levels of depressive symptoms on Beck's Depression Inventory (BDI), capturing correlations between facial appearance and symptoms of depression. These composites were then used in a warping procedure to create a high and low-symptom morph of 20 men and 20 women. That is, a high-depression image and a low-depression image were created for each of the 40 individuals in the set. We used the Scott et al.'s stimulus set to create two new sets of stimuli. The first set comprised the original images cropped to leave a T shaped image containing on the internal facial features; in the second set, the faces were cropped to leave only the eye region, as illustrated in figure 3.1. Therefore, for each stimulus set (internal features or eyes) there were high and low-depression image pairs for each of 20 men and 20 women.

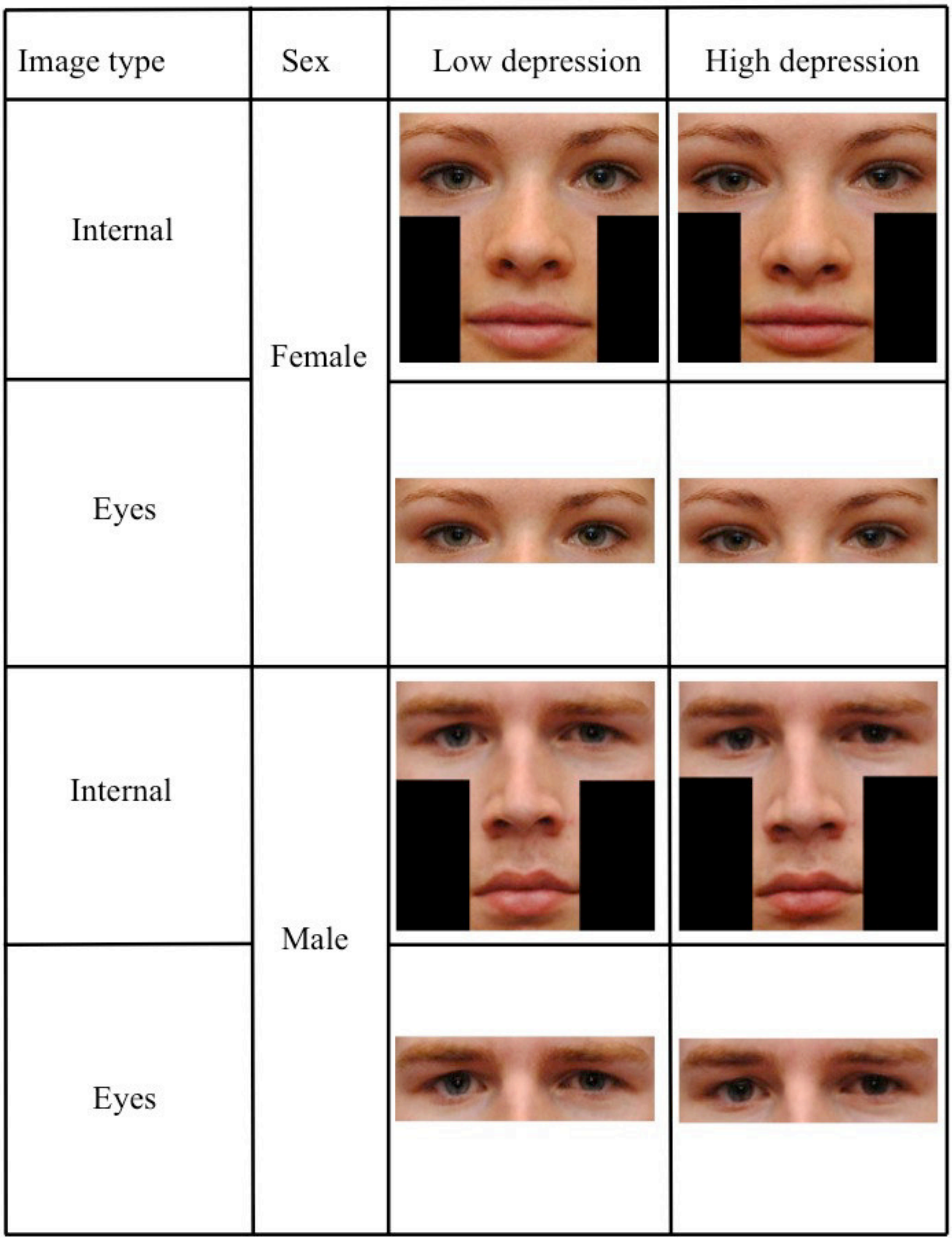


Figure 3.1. Example high and low depressive symptom stimuli reduced to show internal features or just eyes.

Study one - symptom recognition

In this study we assessed observers ability to accurately differentiate between the high and low depressive symptom stimuli when it was reduced to only show internal features and further to show just the eye region. The study was approved by Bangor University's ethical review board.

3.2.2. Observers: Thirty-one (24 female) undergraduate students (Age *M* = 23.44, *SD* = 4.12) were compensated with course credits for their participation.

3.2.3. Procedure: The presentation of stimulus set (internal features or eyes) was blocked and counterbalanced across participants. On each trial, observers were presented with a pair of images, consisting of the high-depression and low-depression image for a single individual. A discrimination statement appeared below the images, asking observers to either choose which image had 'More symptoms of depression' or 'Fewer symptoms of depression'. Observers used the mouse to click on the image of their choice.

Each image pair appeared twice within a block, once with the 'more symptoms' and once with the 'fewer symptoms' discrimination statement. Face pairs were presented in random order within a block. Position of the images was balanced so that position was uncorrelated with the level of depression. Viewing distance was not fixed, and responses were not speeded.

3.2.4. Results

Correct answers were defined as observers clicking on the low-depression image when asked to pick 'Fewer symptoms of depression' or the high-depression image when asked to pick 'More symptoms of depression'. Our main interest was in whether depressive symptoms could be accurately identified in the internal features and the eyes conditions. In fact, observers were significantly better than chance for both conditions: with mean observer accuracy of 58%, 95% CI [55-61%], for internal features, $t(30) = 6.08$, $p = .00001$, $d = 1.12$; and 56%, 95% CI [52-59%], mean accuracy using only the eye region, $t(30) = 3.73$, $p = .0008$, $d = .65$. Accuracy was numerically higher when viewing internal features than eyes alone, although not significantly, $t(30) = 1.05$, $p = .301$, $d = .19$.

We conducted a 2 (source: eyes, internal features) x 2 (stimuli sex) x 2 (participant sex) mixed ANOVA to check whether accuracy was limited to any particular set of conditions. However, we found no significant effects of source, $F(1, 29) = 3.58$, $p = .068$, stimuli sex, $F(1, 29) = 0.001$, $p = .716$, or participant sex, $F(1, 29) = 1.06$, $p = .312$, and no interactions between variables (all *p*'s > .1). Accuracy in identification therefore appears largely comparable across all conditions.

Finally, we compared our results with limited features to Scott et al.'s (2013) original study using full faces. Independent samples *t*- tests found accuracy to be marginally decreased when viewing full faces versus internal features, 61 % vs 58%, $t(93) = 2.01$, $p = .047$, $d = .42$, and significantly decreased between full faces and eyes only, 61% vs 56%, $t(93) = 2.83$, $p = .006$, $d = .59$.

These results show levels of depressive symptoms can be identified at levels above chance merely from the eye (and eyebrow) region. Because accuracy in the eyes-alone and internal-features conditions were significantly lower than in the full-face stimuli, there is likely to be valid signals to depression in the external features of the face including the jaw line or that the cues are more viable when put into context with the eyes.

3.3. Study two - personality attributions

In study one we demonstrated that observers were able to accurately differentiate between facial images of individuals with high and low depression symptom levels, even when the stimuli were reduced to show just the eye region. In experiment 2 we sought to determine if the negative personality traits Scott et al. found to be attributed to individuals with high symptoms levels were also upheld with this impoverished stimuli. The study was approved by Bangor University's ethical review board.

3.3.1. Observers: Thirty-two (28 female) observers (age $M = 19.56$, $SD = 2.14$) were compensated with course credits for their participation.

3.3.2. Procedure: The procedure largely replicated that of study one, with internal and eyes-only stimuli being presented in two counterbalanced blocks. Within each block observers were presented with each warped image pair, side by side, and asked to pick which image they felt best fitted the description given by clicking on it with the mouse. Each pair was presented five times where observers were asked to pick the image which they thought was more: Open ("Who looks more open to new experiences"), Agreeable ("Who looks more friendly and compassionate"), Extraverted ("Who looks more outgoing and energetic"), Conscientious ("Who looks more efficient and organised") and Neurotic ("Who looks more sensitive and nervous"). The frequency with which observers picked the low-depression image to fit the descriptions was recorded.

3.3.3. Results

Internal features: As shown in figure 3.2 observers were largely at chance level when responding to personality attributions based on viewing only the internal features of each face. The two responses which did rise significantly above chance levels were for Openness (58%, 95% CI [54-61%], $t(31) = 4.73$, $p < .001$, $d = .87$) and Extraversion (56%, 96% CI [51-61%], $t(31) = 2.49$, $p = .022$, $d = .44$) in male images.[3]

[3] Please note when results are corrected for multiple comparisons using Bonferroni Sequential corrections (Holm, 1979) only ratings of male Openness retain significance ($p = .0047$).

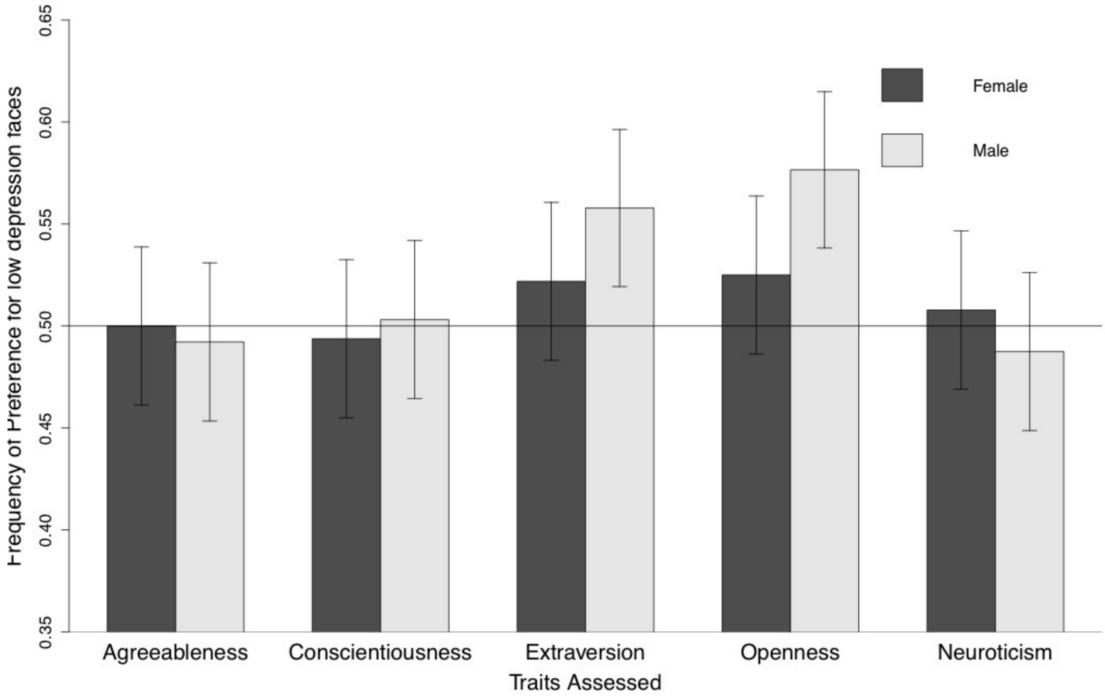


Figure 3.2. Results showing the frequency observers picked the internal features high depression image as having more symptoms of depression, Error bars represent 95% confidence interval of the mean for each trait.

Eyes-only: As shown in figure 3.3 again results were largely around chance levels. Once more the only results to rise significantly above chance levels were responses to male images for Openness (55%, 95% CI [51-60%], $t(31) = 2.33$, $p = .026$, $d = .41$) and Extraversion (55%, 95% CI [51-59%], $t(31) = 2.15$, $p = .04$, $d = .42$). [4]

[4] Please note, when results are corrected for multiple comparisons using Bonferroni sequential corrections they fail to reach significance.

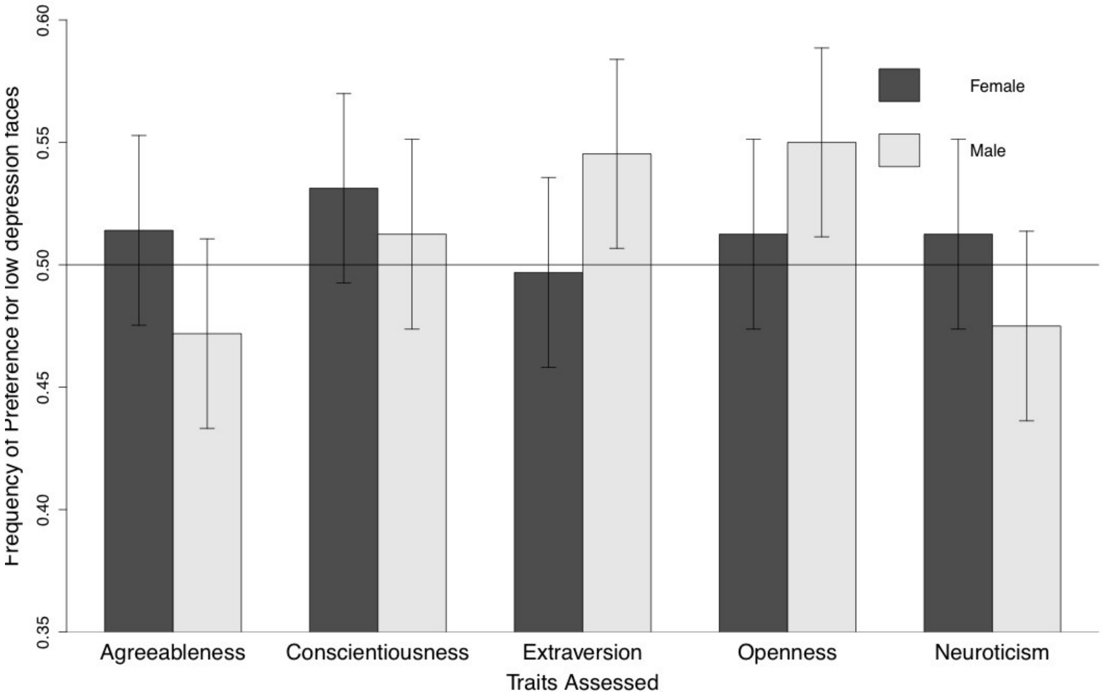


Figure 3.3. Results showing the frequency observers picked the eyes-only high depression image as having more symptoms of depression. Error bars represent 95% confidence interval of the mean for each trait.

Comparisons: Both low depression internal and eyes-only stimuli were perceived to be higher in Openness and Extraversion, however paired samples *t*-tests showed that there was not a significant difference between response rate to the two types of stimuli ($t(31) = 1.13$, $p = .27$, $d = .23$, and $t(31) = 0.45$, $p = .66$, $d = .006$ respectively).

3.4. Discussion

This study sought to identify regions of the face which would be important for the accurate identification of depressive symptoms, as first described by Scott et al. (2013). Using two variations of the stimuli used by Scott et al. The first one occluded the external contour of the face and jawline to leave only a T shaped section showing the internal features of the face: eyes, nose and mouth. The second presented only the region around the eyes. It was found that observers were still able to accurately discriminate between high and low depressive symptoms when presented with just internal features and even just the eyes.

These results shed some light on the possible areas of the face used as cues to depressive symptoms. The drop in accuracy found between the internal features used here and the full faces presented by Scott et al. (2013) suggests that there are cues in external facial features, such as jaw line and muscle tension in the cheeks and forehead, which can aid the accurate identification of depressive symptoms. However, the similar levels of accuracy found for internal features and for eyes stimuli suggests that a majority of the cues within the internal features are found in the region around the eyes. This is in accordance with previous research theorising a link between depression and the personality trait, Neuroticism (Bagby, 1997), alongside additional research suggesting that eyes could be a cue to Neuroticism (Paunonen, Ewan, Earthy, Lefave, & Goldberg, 1999).

Additionally it was sought to determine if the negative personality attributions given to high-depression images found by Scott et al. (2013) would still hold when the stimuli was impoverished. Recall that Scott et al., using full faces, found that composites for high depressive symptoms were both accurately discriminated from low depression composites, but also were given global negative trait attributions. Here there was some dissociation between discrimination accuracy and trait judgments. The occluded faces used here presented enough information for accurate discrimination, but did not trigger the global downgrading found by Scott et al. Therefore while internal features contain much of the information required for discrimination of depressive symptom levels, it appears that the negative downgrading may be made more holistically on the basis of the full face. The ability to accurately detect personality traits from stimuli of a similar style has been previously demonstrated (Kramer & Ward, 2010), suggesting that the lack of attributions found here is not due to a lack of ability. However, it may be reminiscent of the way in which attractiveness judgments have been found to be dependent on holistic judgments rather the combination of face parts (Abbas & Duchaine, 2008). Indeed, the lower levels of attractiveness found by Scott et al. for the high depression faces are based on the full face.

There are limitations to this research, namely the use of a student, rather than clinical sample of faces with depressive symptoms. Although the promising results give a good basis for the study replication with a clinical sample, showing the degree to which depressive symptoms can be identified from the face. Additionally, although the accuracy levels reported are significantly above that expected by chance, results of 56-58% are far from perfect identification. There are two ways to think about this outcome. First, this level of accuracy is not high in absolute terms. It would therefore be foolish to diagnose depression on the basis of the eyes alone, for example. Second, however, is that it is noteworthy that observers could use a stimulus so intentionally impoverished as a static image of just the eyes to make any valid inferences at all about a social trait as important as depressive symptoms. In this respect, the findings therefore highlight the rich range of cues unintentionally expressed in the face but accurately decoded by observers An area of future exploration would assess the part played by observer reactions to these unintentional facial cues in outcomes relating to depression.

The next step

In this chapter I demonstrated that naïve observers were still able to accurately distinguish between individuals with high and low levels of depressive symptoms even when they were only presented with the eye and brow region. In the next chapter I investigated if the ability to detect mental health traits would extend to another common psychiatric disorder, schizotypal personality disorder.

Chapter four: The recognition of schizotypal personality disorder from static facial images

Abstract

Personality disorders are commonly thought of as a representation of a maladaptive personality style, often distinguished using the Big Five personality traits. Schizotypal personality disorder (SPD) is a schizophrenia spectrum disorder which has been categorised by high scores in Neuroticism and low in Extraversion (Saulsman & Page, 2004). Previous research has demonstrated the ability to detect these personality traits from static facial images (Kramer & Ward, 2010). This study aimed to determine if naïve observers are able to utilise these cues to accurately categorise individuals with and without symptoms of SPD. Utilising two samples of participants, two measures of SPD symptoms, and two methods of stimuli creation I demonstrated that observers were consistently able to discriminate between male high and low symptom scorers, findings for female images were found to be more inconsistent. In addition to this high SPD images were found to be continuously rated as having less desirable personality traits, measured by the Big Five personality traits, as well as general perception of social desirability and trustworthiness. The implications for the social perceptions of individuals with SPD these findings present are discussed.

4.1. Introduction

The Five Factor model of personality, consisting of the personality traits Agreeableness, Conscientiousness, Extraversion, Openness and Neuroticism, has been increasingly used as a theoretical alternative to the categorical approach of the DSM. For example, a recent meta-analysis showed a general trend for individuals with mental health disorders to be low in traits of Agreeableness, Conscientiousness and Extraversion, as well as high Neuroticism. Results found to be consistent across clinical and non-clinical samples (Malouff, Thorsteinsoon, & Schutte, 2005).

An area where this approach has taken particular hold is that of personality disorders. Early work in this area, as reviewed by Dyce (1997), was promising, however, agreement was varied as a product of reporting method, analysis and item measurement. More recently conducted research appears to have overcome these criticisms to produce a consistent personality profile of individuals with personality disorders. Using an 'expert' approach Lynam and Widiger (2001) asked researchers with experience of personality disorders to rate how they believed individuals with each of the 10 personality disorders, as defined by the DSM-IV (American Psychiatric Association, 1994), would score on the 30 facets of personality in the NEO-PI-R (Costa & McCrae, 1992). In general they rated individuals with personality disorders to be high in Neurotic traits of anxiousness and self-consciousness, whilst low in Extraverted traits of warmth, gregariousness and positive emotions, the Openness trait of open to new ideas and Conscientious trait of order. Interestingly, although inter-rater reliability was generally found to be satisfactory, one of the lowest agreement levels was found for schizotypal personality disorder (SPD).

Further research provides support for the use of the five-factor model of personality in SPD, such as results from Gurrera et al. (2005). When comparing a community sample (SPD diagnosis determined using SCID-II interviews) to healthy controls, individuals with SPD scored significantly higher for Neuroticism and Openness, but lower for Agreeableness, Conscientiousness and Extraversion. Bagby, Costa, Widiger, Ryder and Marshall (2005) found almost identical results using clinical patients and obtaining measures of the big five personality traits through both self-report and interview. Again, individuals with SPD were found to be high in Neuroticism, and low Agreeableness, Conscientiousness and Extraversion – results consistent across both personality measures. Results for Openness failed to reach significance, however Gurrera et al. (2005) suggest that the significant result for Openness in their study was caused by a sample pool consisting entirely of female participants, creating a possible bias.

Collaborating results from studies assessing the connection between the five factor model and personality disorders from 1990-1998 Saulsman and Page (2005) largely support previous findings, reporting significant ($r > .20$) correlations between high Neuroticism and Low Extraversion and SPD. Results for Agreeableness were in the direction found by other researchers, but failed to reach the pre-determined significance cut off ($r = -.11$). Additionally, using binomial weighted means, Saulsman and Page (2005) estimated that 68% of individuals with SPD would have high Neuroticism scores, 64% low Extraversion, 60% low Agreeableness, 56% low Conscientiousness and a 50/50% split for Openness.

These results suggest that SPD can be thought of in terms of a specific combination of the five-factor model of personality, namely high Neuroticism and low Agreeableness and Extraversion. It is possible that this personality structure would be visible from the face, as recent research has demonstrated the accurate identification of personality traits from static facial images. In early research Penton-Voak, Pound, Little and Perrett (2006) found significant correlations between self-reports of Agreeableness and Extraversion and ratings given to individual's facial photos. Building on this, Little and Perrett (2007) created averaged 'composite' images from facial photographs of high and low scorers on personality traits which were presented to participants. They found observers to be accurate in discriminations of Neuroticism, Agreeableness and Conscientiousness. More recently Kramer and Ward (2010) demonstrated the accurate identification of Agreeableness, Extraversion and Neuroticism from composite facial images.

There is also some evidence for phenotypic differences between SPD and control individuals relating to facial morphology. Often these studies use the Waldrop scale (Waldrop, Pedersen, & Bell, 1968), which lists specific abnormalities and corresponding weighted scores, to gain a standardised score for minor physical abnormalities (MPAs). Schiffman et al. (2002) found individuals with schizophrenia and other schizophrenia spectrum disorders (such as SPD) to have 3.52 times more MPAs than matched controls. Specifically looking at individuals with SPD Weinstein, Digorio, Schiffman, Walker and Bonsall (1999) found similar results, with SPD diagnosed individuals having significantly more MPAs than controls.

This study aimed to determine if observers would be able to accurately distinguish between facial images of individuals scoring high and low on a measure of SPD. As discussed, previous research has demonstrated an increase in MPAs between individuals with SPD and controls, although no research has been conducted to determine if observers are able to draw on these in a diagnostic manner. Additionally it was sought to determine if observers would perceive individuals with SPD to have the personality traits previously associated with the disorder. That is, individuals with SPD have been generally shown to have high Neuroticism, but low Agreeableness and

Extraversion scores, traits previously shown to be identifiable from the face, but do observers perceive individuals with SPD to have these traits from facial images alone? In order to explore these hypotheses facial photographs were collected alongside a measure of SPD from university students, which were then used in two rating experiments. In each of these experiments a different technique was used to combine the images and create stimuli, and a variety of methodologies were used to present these stimuli.

4.2. Method

This study consisted of two phases, during phase one "Photo collection" participant photos and SPD scores were collected, which were then used to create stimuli for three rating experiments in phase two.

4.2.1. Phase One – Photo collection

4.2.1.1.Participants: Two-hundred-and-twenty-one (130 female) participants (age M = 21.65, SD = 5.09) took part in photo shoot B (See appendix B for details) and were paid £6 for their participation.

4.2.1.2. Measure: Participants completed the Schizotypal Personality Questionnaire (SPQ, Raine, 1991). This is a 72 item questionnaire where symptoms are identified using a "Yes/No" criteria with a possible score range of 0-72. Male scores ranged from 1-60 (M = 25.21, SD = 13.36), and females from 0-67 (M = 24.10, SD = 15.09).

4.2.1.3. Procedure: Facial photographs were taken against a light background whilst participants maintained a neutral facial position after removing any make-up and jewellery and tying hair back from their face. Camera height, flash, zoom and distance from participant (2m) was kept constant.

Participants who were aged over 30, non-White and males with beards and moustaches were removed from the sample in order to maintain composite quality. This left a sample of 100 females and 66 males.

The highest (male score $M = 39.72$, $SD = 9.81$, female score $M = 44.50$, $SD = 8.02$) and lowest 18 (male score $M = 11.05$, $SD = 5.72$, female score $M = 5.33$, $SD = 2.82$) scorers were selected. These were then grouped into nine sets of four participants. Over each of these nine sets participants appeared only twice, and no set of four had more than one participant overlap with any other set, i.e. no composite image had two of the same participants in them (see Appendix A for methodology). Photographs of the four participants in each set were then morphed together, resulting in nine high and nine low composite images for each sex, an example of which can be seen in figure 4.1. These were then used in the following two-part experiment.

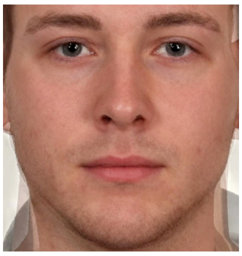
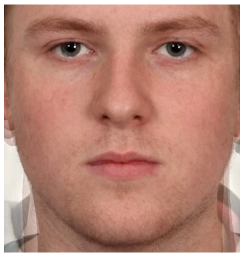
Study	Sex	Low SPQ	High SPQ
One	Male		
	Female		
Two	Male		
	Female		

Figure 4.1. Example stimuli of high and low SPD stimuli. Firstly developed from photographs of individuals scoring high and low on the 72 item SPQ, and secondly from photographs of individuals scoring high and low on the 22 item SPQ-B.

4.2.2. Phase Two

During this phase the stimuli created in phase one were then presented to observers in a two-part study assessing the observer's ability to discriminate between high and low SPD images and their perceptions of the personality traits associated the images. The order of study presentation was counterbalanced across observers.

4.2.2.1. Observers*:* Thirty-five (23 female) observers (age *M* = 20.60, *SD* = 3.99) were compensated with course credits for their participation.

Experiment 1.1 - SPD discrimination

This experiment assessed observer's ability to accurately discriminate between facial images of individuals scoring high and low for a measure of schizotypal personality disorder.

4.3.1. Procedure: All possible high/low composite combinations for each sex were presented to observers side by side on a computer screen, resulting in 162 trials (81 for each sex). Observers were presented with each composite pair side by side (position counterbalanced) on a computer screen below a statement asking them to pick the image which had 'More symptoms of schizotypal personality disorder'. Observers indicated their decision by clicking the appropriate image with the mouse, which initiated the next trial. Response times were not limited and viewing distance was not set. Observers saw each pair in a random trial order.

4.4 Results

A 2 (observer sex) x 2 (image sex) ANOVA showed significant effects of image sex, $F(1,33) = 33.68$, $p = .00002$, and observer sex, $F(1,33) = 5.08$, $p = .031$, thus results were separate by both variables. Male observers picked the male high SPQ image as having more symptoms of SPD in 76%, 95% CI [61-90%], of trials, significantly more often than expected by chance, $t(11) = 3.84$, $p = .0027$, $d = 1.1$. However, results for male judgements of female images did not reach significance with the high SPQ image only being picked to have more SPD symptoms 48%, 95% CI [38-58%] of the time, $t(11) = 0.43$, $p = .677$, $d = .12$. Results for female observers varied slightly, with the male high SPQ image being picked as having more symptoms in 82%, 95% CI [77-86%], of trials, significantly above chance levels, $t(22) = 14.77$, $p <$

.000001, *d* = 3.08. When observing female SPQ images female observers picked the high image as having more symptoms in 58%, 95% CI [51-66%], of trials, an accuracy that reaches marginal significance, $t(22) = 2.37$, $p = .027$, $d = .49$.

Experiment 1.2– Perceptions of personality

4.5.1.Procedure*:* In this experiment each composite image was presented individually alongside a discrimination statement and a seven point Likert scale. Observers were asked to rate each composite image for Agreeableness ‘How agreeable is the person? e.g. more co-operative, more kind and considerate, more generally trusting, more helpful and unselfish’, Conscientiousness ‘How conscientious is this person? e.g. more thorough, a more reliable worker, perseveres more with tasks, more likely to make plans and see them through’, Extraversion ‘How extraverted is this person? e.g. more talkative, more enthusiastic, more assertive, more outgoing and sociable’, Neuroticism ‘How neurotic is this person? e.g. more depressed and blue, worries more, gets nervous more easily, more moody’, Openness ‘How open is this person? e.g. values more artistic, aesthetic experiences, more likely to reflect and play with ideas, more curious about many things, ore inventive’, attractiveness ‘How attractive is this person?’ and trustworthiness ‘How trustworthy is this person?’. This resulted in 252 trials, the order of which was randomised and participants were asked to indicate their choice by clicking on the appropriate number with the mouse.

4.6. Results

A 7 (trait) x 2 (observer sex) x 2 (image sex) x 2 (symptom severity) Factorial ANOVA showed a significant effect of image sex, $F(1,33) = 81.43$, $p < .000001$, and as expected trait, $F(6,198) = 9.31$, $p < .000001$, and symptom severity, $F(1,33) = 83.18$, $p < .000001$, but there was no significant effect of observer sex $F(1,33) = 0.01$, $p = .92$. For this reason results were aggregated over observer sex but split over other factors.

As shown in Table 4.2 high SPQ male faces were rated as significantly lower than low SPQ faces on traits of Agreeableness, $t(33) = 7.38$, $p < .00001$, $d = 1.05$, Conscientiousness, $t(33) = 10.05$, $p < .00001$, $d = 1.79$, Extraversion, $t(33) = 10.80$, $p < .00001$, $d = 1.85$, Openness, $t(33) = 8.40$, $p < .00001$, $d = 1.44$, Attractiveness, $t(33) = 9.50$, $p < .00001$, d = 1.63, and Trustworthiness, $t(33) = 7.23$, $p < .00001$, $d = 1.23$, but higher in Neuroticism, $t(33) = 6.21$, $p < .00001$, $d = 1.07$. Results for female images were less clear cut, as the high SPD image was perceived to be significantly higher in both Agreeableness, $t(33) = 3.85$, $p = .0005$, $d = .67$, and Trustworthiness $t(33) = 2.11$, $p = .04$, $d = .36$. [5]

Table 4.2. Table detailing the average score given to high and low scorers of SPD for each personality trait, on a scale ranging from 1(very low) to 7 (very high).

	Stimuli sex	High SPD	Low SPD	95% CI	*d*
Agreeableness	Male	3.24*	4.32*	-1.38 to -0.79	1.05
	Female	4.70*	4.35*	.166 to .54	0.66
Conscientiousness	Male	3.23*	4.46*	-1.43 to - 0.95	1.79
	Female	4.71	4.71	-0.16 to 0.15	0.00
Extraversion	Male	3.34*	4.56*	-1.45 to - 0.99	1.85
	Female	4.42	4.34	-0.07 to 0.23	0.05
Openness	Male	3.03*	4.35*	-1.64 to - 1.0	1.44
	Female	4.56	4.51	-0.21 to 0.29	0.07
Neuroticism	Male	4.38*	3.28*	0.74 to 1.46	1.07
	Female	3.46	3.48	-0.34 to 0.31	0.02
Attractiveness	Male	3.00*	4.33*	-1.62 to -1.04	1.63
	Female	4.42	4.32	-0.11 to 0.30	0.17
Trustworthiness	Male	3.24*	4.33*	-1.40 to -0.79	1.23
	Female	4.67*	4.49*	0.0071 to 0.36	0.36

$*p < .05$

Experiment 2

In Experiment 1 we demonstrated that observers were able to accurately discriminate between high and low male SPD images, and attributed more negative

[5] Please note when correcting for multiple comparisons results for female Trustworthiness fail to reach significance ($p = .28$).

personality traits to those images. Results for females were less clear-cut, with accurate discrimination only being demonstrated by female observers and personality attributions favouring high SPD images. In this Experiment we sought to replicate Experiment 1 with a new stimuli set, created from an abbreviated version of the SPQ, rated by new observers.

4.7. Phase one - Stimuli creation

As in Experiment 1 stimuli were created from facial photographs collected during a university photo-shoot. The procedure largely replicates that of Experiment 1, with any differences noted below.

4.7.1.Participants*:* Two hundred and forty (141 females) Bangor university students (Age *M* = 21.69, *SD* = 5.19) took part in photo shoot B (See appendix B for details) and paid £6 for their participation.

4.7.2. Measure: Participants completed the Schizotypal Personality Questionnaire - Brief (SPQ-B, Raine & Benishay, 1995). This is a reduced version of the SPQ used in study 1, consisting of only 22 of the SPQ's 72 questions. As with the SPQ participants indicate whether they agree or disagree with each statement given, giving a possible total score of 22. Female scores ranged from 0-22 and male scores ranged from 1-21.

4.7.3. Procedure: The photograph collection procedure replicated that of study 1, and once again the highest (Female *M* = 15.89, *SD* = 2.12, Male *M* = 12.82, *SD* = 2.69) and lowest (Female *M* = 1.28, *SD* = 0.90, Male *M* = 3.39, *SD* = 1.70) 18 scorers were selected and using the balanced design described previously, were used to create nine high and nine low composite images for each sex. Examples shown in figure 4.1.

4.8. Phase Two

The experimental procedure of Experiment 1 was matched, with a the new stimuli developed in Phase One and a new set of observers.

4.8.1. Observers*:* Twenty-eight (23 female) observers (Age *M* = 19.86, *SD* = 2.18) were compensated with course credits for their participation.

4.8.2. Experiment 2.1 - SPD discrimination

4.8.2.1 Procedure*:* This experiment replicated experiment 1.1.

4.9. Results

A 2 (Image sex) x 2 (Observer sex) Mixed ANOVA showed a significant effect of image sex, $F(1, 26) = 13.02$, $p = .001$, but no significant effects of observer sex $F(1, 26) = 1.25$, $p = .27$ and an non-significant interaction between the two, $F(1, 26) = 0.03$, $p = .87$. For this reason results were split by image sex, but collapsed over observer sex. Observers were accurate in their discrimination of Male SPD stimuli in 59%, 95% CI [55-62%], of trials, a rate significantly above that expected by chance, $t(27) = 5.83$, $p < .00001$, and in 72%, 95% CI [66-77%], of trials with Female images, again significantly above chance rates $t(27) = 7.96$, $p < .000001$.

4.10. Experiment 2.2 - Personality attributions

This experiment replicated experiment 1.2, with any differences noted below.

4.10.1.Procedure*:* In this experiment observers were presented with each composite image three times and asked to rate them on a 7-point Likert scale for femininity (“How feminine is this person?”), attractiveness (“How attractive is this person?”) and social desirability (“How socially desirable is this person?”).

4.11. Results

A 3 (Trait) x 2 (Observer sex) x 2 (Image sex) x 2 (Symptom severity) mixed ANOVA showed a significant effect of trait, $F(2, 52) = 8.10$, $p = .001$, symptom severity $F(1, 26) = 47.75$, $p < .001$, and image sex, $F(1, 26) = 131.72$, $p < .001$ but not Observer sex, $F(1, 26) = 0.81$, $p = .38$. For this reason results were collapsed over observer sex but separated by all over variables.

As shown in Table 4.3, observers attributed positive personality characteristics to low SPD images. Observers perceived low SPD female images to be more feminine, $t(27) = 7.76$, $p < .000001$, $d = 0.12$, Attractive, $t(27) = 10.81$, $p < .000001$, $d = 2.23$, and socially desirable, $t(27) = 6.21$, $p < .00001$, $d = 1.17$, and male images to be marginally more attractive, $t(27) = 2.07$, $p = .048$, $d = 0.36$, and socially desirable, $t(27) = 2.14$, $p = .04$, $d = 0.41$, but less feminine, $t(27) = 2.25$, $p = .03$, $d = 0.42$. [6]

	Stimuli sex	High SPD	Low SPD	CI	*d*
Attractiveness	Male	3.12*	3.32*	-0.39 to - 0.002	0.36
	Female	3.38*	4.51*	-1.35 to -0.92	2.23
Femininity	Male	2.26*	2.10*	0.01 to 0.31	0.42
	Female	4.37*	5.30*	-1.17 to -0.68	0.12
Social desirability	Male	3.90*	4.06*	-0.31 to - 0.007	0.41
	Female	4.04*	4.81*	-1.02 to - 0.52	1.17

Table 4.3. Table detailing the ratings given to high and low SPD scorers for attractiveness, femininity and social desirability. * $p < .05$.

4.12. Discussion

In this study we created two sets of stimuli representing individuals scoring high and low on measures of schizotypal personality disorder, which were then presented to two sets of observers in two studies aimed to determine if individuals can accurately

[6] Please note when correcting for multiple comparisons results for femininity ($p = .08$) and social desirability ($p = .09$) in females reduced to chance level.

discriminate between high and low scorers and if there are personality traits consistently attributed to individuals with high SPD symptoms.

In experiment 1.1 both male and female observers were found to be highly accurate in the discrimination of male SPD images, but only females performed above chance levels when responding to female images. Similar results were found in experiment 1.2, where observers were accurate in discriminations of both male and female SPD images.

One possibility for the inconsistent findings with female images is the gender bias in the diagnosis in SPD, where men are more likely to be diagnosed (Boggs et al., 2009) and have high SPQ scores (Bedwell & Donnelly, 2005). Additionally, females with SPD have been found to have a less severe cognitive impairment (Voglmaier et al., 2005). However, in this sample male and female scores were generally similar. Another possibility is the dimension of SPD used as a cue, for example males have been found to have higher scores than women on the disorganisation subscale, as well as individual questions of no close friends and odd behaviour (Fossati, Raine, Carretta, Leonardi, & Maffei, 2003), traits which would be important in social interaction. Similar results were obtained by Raine (1992), who found women to be higher on the positive SPQ subscale, whereas males were found to be higher on the negative subscale.

Additionally, in experiment 1.2 high male SPD images were perceived to be low in Agreeableness, Conscientiousness, Extraversion, Openness, trustworthiness and attractiveness and high in Neuroticism. Previous research has demonstrated a personality type associated with SPD, namely high Neuroticism and low Agreeableness and Extraversion (Saulsman & Page, 2005). These traits were mirrored in the perceptions of high SPD male images in experiment 2.2. These results not only reflect a degree of accuracy, in that observers rated male SPD images to have personality traits previously found to be associated with the disorder, but they also give insight into how observers perceive males with SPD. These perceptions could have a further influence on the social perceptions of males with SPD as personality traits are important factors in determining friendship and social inclusion. For example Berry et al. (2000) found individuals with high Extraversion scores to have closer friendships, friends of highly Agreeable individuals reported less irritation and friendships with highly Open individuals had fewer reported conflicts. In addition Selfhout et al. (2010) found that individuals scoring highly for Agreeableness had more new acquaintances that wanted to form friendships with them. This literature suggests that individuals high in these personality traits (Extraversion, Agreeableness and Openness) are perceived to be desirable in friendships, traits which high male SPD images were consistently rated as lower in. This could have social consequences, in that males with SPD may have tendencies towards feelings of paranoia and social exclusion, which are then exacerbated by social buffing from others due to their involuntary signalling of social undesirable personality traits.

Similar perceptions of social desirability in clinical syndromes were found by Scott, Kramer, Jones and Ward (2013), where observers rated facial images of individuals with high depressive symptoms to be less Agreeable, Conscientious, Extraverted, Emotionally stable (the reverse of Neurotic) and Open. These results not only mirror the findings found with males in the present study, but also the personality traits found to be associated with clinical syndromes in general by Malouff, Thorsteinsoon and Shutte (2005). In their meta-analysis Malouff et al. (2005) demonstrated a general personality type for individuals with clinical disorders; high Neuroticism alongside Low Agreeableness, Conscientiousness and Extraversion. It is interesting to note that this personality make up has been shown to be cued in faces of individuals with depressive symptoms and SPD, although it must be noted that findings were only valid for males with SPD.

Perceptions of social desirability found for male images were not replicated for females in experiment 1.2. Interestingly, Gurrea et al. (2005) found a negative correlation between self reported Extraversion scores and negative SPD symptoms. The accurate identification of Extraversion from facial images has been consistently reported (Kramer & Ward, 2010; Peton-Voak et al., 2006). This could be a possible reason for the lack of accuracy in female faces, if cues of extraversion were used to help determine symptoms (or vice versa) and females were lower scoring on these traits the cue may be diminished. This may also account for the contrary significant findings in experiment 2.2, where using a much more general description of social desirability, both male and female high SPD scorers were found to be rated as less socially desirable than low scorers, although the significance of the findings in females was reduced when corrections for multiple comparisons were applied. Interestingly this experiment utilised stimuli in which observers were able to accurately discriminate between high and low scorers in both sexes. Although these findings suggest that perceptions of personality may be utilised as cues in these discriminations the methodology does not allow for a causal answer. Further research, utilising a Likert style methodology and regression analysis, should be used to determine if observers are using perceptions of personality traits to accurately discriminate between high and low scorers.

This study is limited in its generalisability due to the use of a student sample, although it produced promising results a follow up study with a clinical sample would be required to determine the extend to which cues of SPD from the face are.

The next step

In this chapter I found participants to be able to accurately discriminate between high and low scorers of symptoms of SPD, although findings for female images were less straightforward. Once more favourable personality traits were attributed to high scorers. In the next chapter I aimed to determine if these findings would be replicated in another common personality disorder, borderline personality disorder.

Chapter five: Personality attributions and borderline personality disorder symptomology

Abstract

There has been a wide array of research suggesting that borderline personality disorder (BPD) would be better categorised in terms of a distinct pattern of personality traits. Specifically, Trull, Widiger, Lynam and Costa (2003) found symptoms to be highly correlated with high Neuroticism and low Agreeableness and Conscientiousness scores. Previous research has demonstrated the ability to detect personality traits from static facial images (e.g. Kramer & Ward, 2010). This study aimed to determine if this ability could be utilised in a novel way, namely to accurately categorise individuals with and without symptoms of BPD. Stimuli were created from the highest and lowest 18 scorers of the borderline symptom list, and presented to participants in a two-alternative-forced-choice style study. It was demonstrated that naïve observers were able to accurately discriminate between high and low female scorers, but performed at chance levels for male images. In addition negative personality traits were attributed to high BPD scorers, in a similar style found by Trull and Colleagues (2003). These have potential social implications for individuals with BPD, as the negative personality traits attributed to high scorers have been shown to be important in social relationships.

5.1. Introduction

Historically, the DSM criteria for borderline personality disorder (BPD) was broad, with individuals only requiring five of the nine criteria to achieve a diagnosis (American Psychiatric Association, 1994), resulting in 151 possible symptom combinations (Skodol et al., 2002). Although a much more stringent criteria is now utilised in the DSM-5 (American Psychiatric Association, 2013), there has been a wealth of research into possible alternative models aiming to overcome the issues of heterogeneity and co-morbidity associated with the DSM diagnoses since their introduction into the DSM-III. This literature will be reviewed in order to gain an insight into the potential implications of personality structure in the perceptions of BPD.

One alternative model is the five-factor model of personality; here personality traits are on a continuum from normal to abnormal functioning. The big five personality factors, proposed by McCrae and Costa (1985), consist of Agreeableness, Conscientiousness, Extraversion, Openness and Neuroticism. In 1995 Costa and McCrae further broke these traits down into six facets each, resulting in 30 facets of personality overall. The five factor model was first researched as an alternative to the categorical, DSM, model by Wiggins and Pincus in 1989. Looking at personality disorders in general, they conducted a factor analysis of self-report measures of personality disorders and the big five personality traits in a student sample. BPD was found to load positively onto the Neurotic factor, suggesting higher Neuroticism scores were associated with higher levels of BPD symptomatology.

These findings were followed up by Costa and McCrae (1990), again using a non-clinical sample, they found a positive correlation between the BPD subscale of the Minnesota Multiphasic Personality Inventory Personality disorder scale (MMPI-PD, Morey, Waugh, & Bashfield, 1985) and Neuroticism and Extraversion, and negatively correlated scores with Agreeableness and Conscientiousness, as measured by the NEO-PI (Costa & McCrae, 1985). These were followed by the first clinical study, which utilised both diagnostic interview and self-report measures of BPD (Trull, 1992). Individuals diagnosed by clinical interview as having BPD were also found to have high levels of Neuroticism and low levels of Agreeableness and Conscientiousness, as determined by the NEO-PI. Participants also completed two self-report measures of personality disorders, high scorers for BPD on the MMPI-PD were found to also have high Neuroticism and low Agreeableness scores. Individuals scoring highly on the Personality Diagnostic Questionnaire-Revised (PDQ-R, Hyler, Skodol, Oldham, Kellman, & Doidge, 1992) were shown to be highly Neurotic and Open, but low in Agreeableness.

Following the example of Trull (1992) comparing results with clinical interviews and self-report measures, Soldz, Budman, Demby and Merry (1993) assessed individuals BPD symptomatology using the Personality Disorder Examination (PDE interview) and the Milon Clinical Multiaxial Inventory II (Millon, 1987), alongside the 50-Bipolar-Self-Rating-Scales (Goldberg, 1992) to measure the Big Five. Using the PDE interview BPD symptoms were positively correlated with Neuroticism and Openness. When using the self-report measure symptoms were also positively correlated with Neuroticism, but also had negative correlations with Agreeableness and Conscientiousness.

Morey and colleagues (2002) compared clinical samples of individuals with a range of personality disorders with a community sample to determine if there was a deviation from normal traits both between clinical and control groups, and within clinical groups. Generally individuals with personality disorders were differentiated from the community sample by higher scores in Neuroticism alongside below average scores in Agreeableness, Conscientiousness and Extraversion. For BPD specifically, patients were found to have Neuroticism scores 2.49 standard deviations higher than the community sample as well as Conscientiousness scores 1.64, Agreeableness 0.82. Extraversion 0.81 and Openness 0.51 standard deviations lower than community controls. Although differences in NEO personality inventory revised (Costa & McCrae, 1992) scores could differentiate personality disorder groups from controls, they were limited in ability to determine between different types of personality disorder, suggesting that disorders represent diverse interactions among traits.

Using a slightly different methodology Trull, Widiger and Burr (2001) assessed the ability of the Structured Interview for the Five-Factor Model (Trull & Widiger, 1997) to predict the symptoms of BPD, assessed by the personality diagnostic questionnaire revised. When the two measures were correlated high symptoms BPD were associated with high scores for Neuroticism and Openness and low scores for Extraversion and Conscientiousness. When the big five and their lower order traits were placed into a regression model, Neuroticism facets of Anxiety and Hostility alongside the Conscientiousness facet of low Deliberation were found to significantly predict high BPD scores.

Once more using the full 30 facets of the Big Five, Lynam and Widiger (2001) sought to determine if clinicians were able to detail a particular combination of personality traits associated with BPD. Experts were asked to detail which personality facets they believed would be outside the normal range in individuals with a range of personality disorders. The 24 individuals who took part has good general agreement for individuals with BPD to be characterised by high general Neuroticism, specifically with traits of anxiousness, angry hostility, depressiveness, impulsivity and vulnerability, alongside high scores in the Openness trait of feeling actions, and low scores in the Conscientious trait of deliberation.

Recently Trull, Widiger, Lynam and Costa (2005) reviewed studies into the association between BPD and the five-factor model. This review supported the notion that BPD is a maladaptive variant of the five-factor model traits; specifically that it correlates with high Neuroticism and low Agreeableness and Conscientiousness scores. In a similar experiment Samuel and Widiger (2008) conducted a meta-analysis of studies assessing personality disorders and the five-factor model, breaking the five factors down to their 30 facets. Overall BPD was found to be strongly correlated with all facets of Neuroticism and negatively correlated with the Agreeableness facets of trust, straightforwardness and compliance, the Conscientious facets of competence, dutifulness, self-discipline and deliberation and the Extraversion facets of warmth and positive emotion.

As reviewed there has been a wealth of research into the association between BPD and the big five personality traits, utilising many different diagnostic questionnaires and a number of diagnostic interviews. Despite the heterogeneity in study procedure there appears to be a continuous personality trait theme running through the research, as summed up by Trull et al. (2008) and Samuel and Widiger (2008), namely high scores in all Neuroticism traits and low scores in aspects of Agreeableness and Conscientiousness. This study aimed to assess whether this combination of personality traits associated with BPD will aid observers to detect borderline personality disorder traits from facial images. There is a wealth of research which has assessed the ability for individuals to be able to accurately discriminate personality traits from static facial images (Kramer & Ward, 2010; Little & Perrett, 2007; Penton-Voak, Pound, Little, & Perrett, 2006; Shevlin, Walker, Davies, Banyard,

& Lewis, 2003), suggesting that the personality traits associated with BPD may aid in the accurate discrimination between facial images of individuals scoring high and low for BPD symptoms. Borderline personality disorder is not the only disorder assessed in terms of the five-factor model, it has been utilised to explain symptoms in a wide range of personality disorder including schizotypal personality disorder (SPD). Indeed, in chapter 4 I assessed the ability to detect SPD from facial images, which showed that observers were able to accurate determine between those scoring high and low for SPD symptoms and also attributed the personality traits associated with the five-factor model to the high SPD images. However, there were not any significant results with female images. The current study seeks to replicate the methodology of the SPD study, to determine if the findings of the five-factor model are upheld and if results for a disorder with a high female bias will yield significant results with female images.

5.2. Method

This was a two-phase experiment; in phase one participant's photographs were collected and used to create facial stimuli. This was then presented to observers in a rating experiment in phase two.

5.2.1. Phase one - Stimuli creation

5.2.1.1. Participants: Two-hundred-and-forty (141 females) participants (Age *M* = 21.69, *SD* = 5.19) took part in photo shoot C (See appendix B for details) and were paid £6 for their participation.

5.2.1.2. Measure*:* Borderline personality disorder symptoms were measured using the 23-item Borderline Symptom list (BSL-23, Wolf et al., 2009). In this questionnaire participants are presented with 23 problems for which they indicate how much they suffered from over the previous week on a scale of not at all (0) to very strongly (4). Overall score is computed by summing all items resulting in a top possible score of 92. Female scores ranged from 0 - 78, males ranged from 0 - 65.

5.2.1.3. Procedure: Participants were asked to remove all make-up and jewellery, tie back their hair and adopt a neutral facial expression. Photographs were taken against a light background from a distance of 2m, with camera flash, height and zoom kept constant. To maintain consistency with stimuli individuals aged over 30, who did not report themselves to be white British and men with beards were removed.

The highest (female: $M = 45.72$, $SD = 19.05$, male: $M = 19.89$, $SD = 2.39$) and lowest 18 (female: $M = 1.89$, $SD = 1.28$, male: $M = 7.26$, $SD = 2.12$) participants were selected. A balanced design (Appendix A) was used to create nine groups of four participants for high and low images of each sex. Each participant was not included in more than two groups, and there was no more than two individuals the same in any one group. The photographs of the members of each group were then averaged together using the Jpsychomorph software to create composite images, examples of which can be found in figure 5.1.

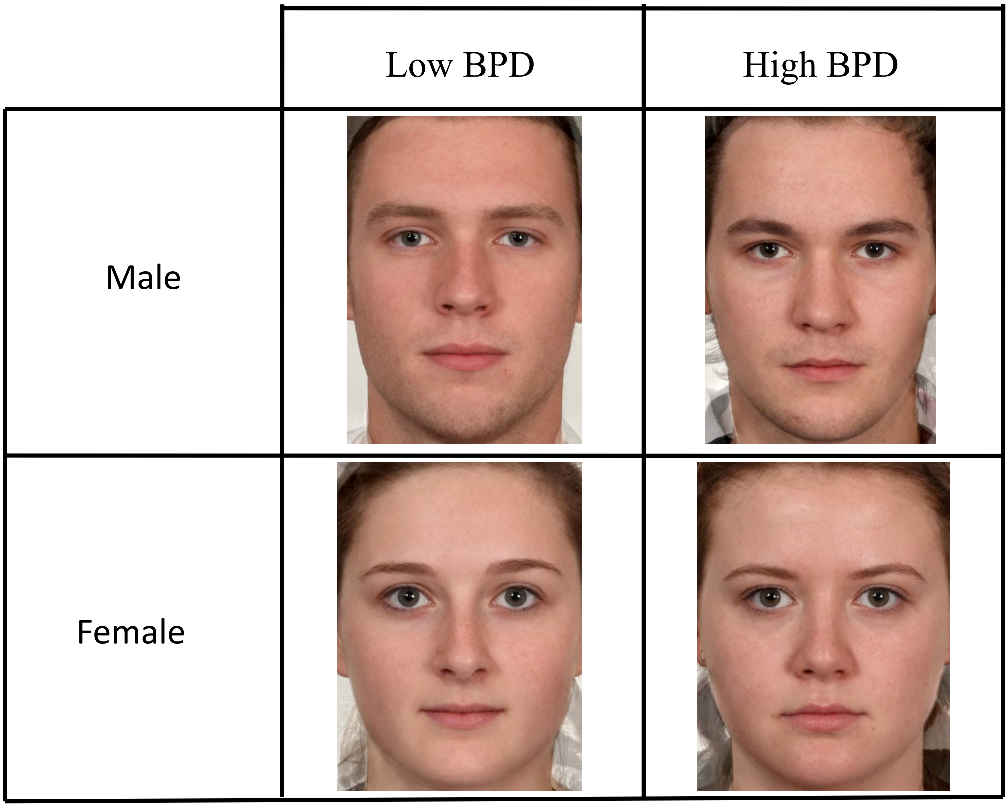


Figure 5.1. Stimuli examples of composite images created from high and low symptom scorers for BPD, each created from photographs of four individuals.

5.2.2. Phase Two - Observer impressions

In this phase the stimuli created in phase one were presented to observers in a two-part experiment. In part one observers were shown pairs of images and asked to determine which one they believed had more symptoms of BPD, and in part two they were asked to rate each face for personality traits.

5.2.2.1.Observers: Twenty-nine (24 females) Bangor University students (age M = 21.00, SD = 5.11) were compensated with course credits for their participation

5.2.2.2. Procedure: Observers were presented with the stimuli in a two-part study, the presentation order of which was counterbalanced across participants.

Accuracy: This was a two-alternative-forced-choice study, used to determine how accurately observers could differentiate between high and low BPD stimuli.

Observers were presented with one high and one low BPD symptom image side by side and asked to pick which image had ‘More symptoms of Borderline Personality Disorder’ by clicking on the image with the mouse. All possible combinations of the nine high and nine low BPD images for each sex were presented, resulting in 162 trials (81 for each sex).

Personality attributions: In this part observers impressions of the stimuli was measured, by rating each image on a seven point Likert scale for each of the Big Five personality traits and attractiveness. These traits were presented in six blocks, the order of which was randomised for each observer. Each block contained 36 trials (one for each image), resulting in 216 trials overall. On each trial the observer was presented with an individual image below a Likert Scale ranging from 1 (very low) to 7 (very high) and trait description, and were asked to use the mouse to click the number they thought best represented the face.

5.3. Results

Accuracy: Results were coded as correct when the high BPD image was selected as having more symptoms of BPD. A 2(Sex of Image) x 2(Sex of observer) ANOVA showed no significant effects of observer sex, $F(1, 27) = 1.04$, $p = .317$, image sex $F(1, 27) = 1.39$, $p = .249$ or an interaction between them $F(1, 27) = 0.01$, $p = .904$. When results were collapsed over these variables it was found observers were able to accurately discriminate between high and low levels of BPD in 54%, 95% CI [52 - 58%] of trials, a result significantly above chance levels, $t(28) = 3.14$, $p = .003$, $d = .46$. However as this accuracy was low further tests were undertaken to determine if there was a numerical, if not statistically significant, difference in accuracy between discriminations in male and female images.

As shown in figure 5.2 observers were found to be accurate in 57%, 95% CI [53 - 60%], of trials with female images, a result significantly above chance levels, $t(28) = 4.32$, $p = .0002$, $d = .77$. However they were only accurate in 52%, 95% CI [48 - 58%], of male images, a result not above that expected by chance, $t(28) = 1.12$, $p = .27$, $d = .14$.

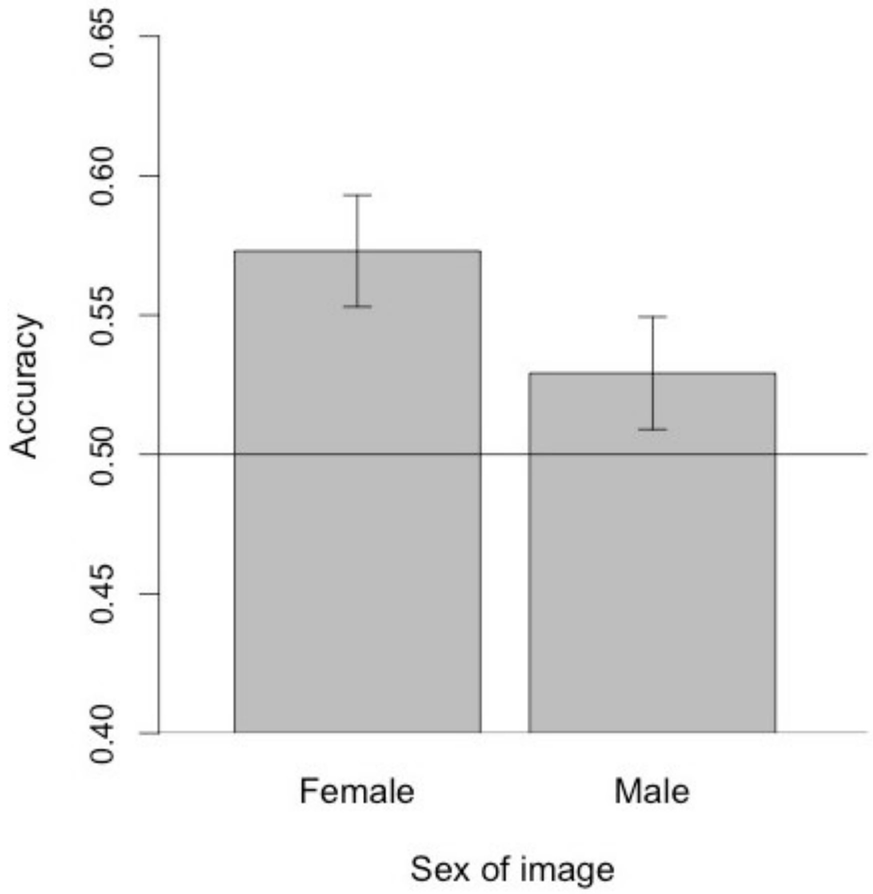


Figure 5.2. Figure showing the accuracy observers obtained when rating high and low symptom images for symptoms of BPD. Error bars represent 95% confidence interval of the mean.

Personality attributions: A 6 (Trait) x 2 (Sex of observer) x 2 (Sex of image) ANOVA showed no significant effect of observer sex, $F(1, 27) = 0.84$, p - .77, but significant effects of image sex, $F(1, 27) = 22.92$, $p < .0001$, and trait, $F(5, 135) = 6.42$, $p < .0001$, so results were collapsed over observer sex and separated over other variables. Likert ratings for each trait given to each of the nine high and nine low male and female images were averaged, the results of which are detailed in table 5.3. Paired samples *t*-tests were used to determine if there was a significant difference between the scores attributed to high and low images. High BPD male images were perceived to be significantly less Agreeable, $t(28) = 3.33$, $p = .0024$, $d = 0.62$, Extraverted, $t(28) = 2.10$, $p = .0446$, $d = 0.37$, and Attractive, $t(28) = 2.44$, $p = .02$, $d = 0.46$. High BPD female images were also perceived to be less Agreeable, $t(28) = 2.59$, $p = .0149$, $d = 0.48$, and Extraverted, $t(28) = 3.28$, $p = .0028$, $d = 0.61$, as well as less Open, $t(28) = 3.37$, $p = .0022$, $d = 0.61$. [7]

Table 5.3. Table detailing Likert scale ratings for personality traits for high and low BPD symptom scorers.

Trait	Image sex	High BPD	Low BPD	CI	*d*
Agreeableness	Male	3.70*	4.08*	-0.61 to -0.15	0.62
	Female	4.39*	4.70*	-0.56 to -0.07	0.48
Conscientiousness	Male	4.13	4.27	-0.36 to 0.09	0.24
	Female	4.57	4.71	-0.30 to 0.03	0.33
Extraversion	Male	3.90*	4.08*	-0.37 to -0.005	0.37
	Female	4.00*	4.29*	-0.47 to -0.11	0.61
Neuroticism	Male	3.82	3.66	-0.05 to 0.44	0.29
	Female	3.95	3.75	-0.05 to 0.38	0.31
Openness	Male	3.97	4.10	-0.37 to 0.11	0.21
	Female	4.20*	4.45*	-0.41 to - 0.10	0.61
Attractiveness	Male	3.47*	3.73*	-0.47 to - 0.04	0.46
	Female	4.39	4.24	-0.13 to 0.44	0.20

* $p < .05$

[7] Please not when correcting for multiple comparisons results for Extraversion ($p = .312$) and Attractiveness ($p = .16$) in males, and Agreeableness ($p = .13$) and Openness ($p = .26$) in females fail to reach significance levels.

5.4. Discussion

In this study facial stimuli was created from images of individuals scoring high and low on a measure of Borderline Personality Disorder symptoms and presented them to observers in a two-part study. In part one it was demonstrated that when results were collapsed over stimuli sex then naïve observers were able to accurately determine between high and low BPD scorers. However further analysis determined that this accuracy was mainly found in response to female, not male, stimuli. In part two it was demonstrated that high BPD images of both sexes were perceived to be less Agreeable and Extraverted, that high BPD women were also seen to be less Open and high BPD men to be less attractive than low images. However, if corrected for multiple comparisons only results for Extraversion in females and attractiveness in males retained significance.

The findings that high BPD images of both sexes are perceived to be lower (even if not stringently significant) in Agreeableness and Extraversion is in line with previous research showing individuals with BPD to have low levels of these traits (Costa & McCrae, 1990; Kendler, Myers & Reichborn-Kjennerud, 2011; Morey et al., 2002; Trull, 1992; Trull et al., 2001; Samuel & Widiger, 2008; Soldz et al., 1993). Interestingly, high BPD females were found to be perceived to be lower in Openness, whereas previous research has found them to be higher (Lynam & Widiger, 2001; Trull, 1992; Trull, Widiger & Burr, 2001 & Solz et al., 1993). The most surprising finding is that although high BPD images were generally rated as more Neurotic, this result did not reach significance. In personality research BPD has been consistently highly correlated with Neuroticism, for example Morey and colleagues (2002) found individuals with BPD's Neuroticism scores to be over two standard deviations higher than controls. Indeed, as part of the stimuli collection participants also completed the mini-IPIP personality inventory (Donnellan, Oswald, Baird, & Lucas, 2006) and BPD scores were significantly correlated with Neuroticism in both males, $r = .54$, $p < .0001$, and females, $r = .39$, $p < .0001$.

Observers have been consistently shown to be able to accurately detect Neuroticism from composite facial images (Jones, Kramer, & Ward, 2012; Kramer & Ward, 2010; Little & Perrett, 2007) so it is unlikely that the reason for the non-significant ratings is an inability to detect Neuroticism. Indeed in much of this previous research individuals high in symptoms of neuropsychiatric disorders, such as depression and schizotypal personality disorder, when observers are able to accurately determine between facial images of high and low symptom level they also perceive the high image to be higher in Neuroticism. A key to the non-significant findings may lie in the wide variability with Neuroticism ratings, reducing the statistical significance but still resulting in a noticeably higher score attributed to individuals with high BPD symptoms. This variability could be attributed to an outlier within the composite set, in that one high BPD composite image was rated, on average, 2 S.D lower in Neuroticism than the others.

The ability to determine between high and low BPD female images was not replicated for males, even though there was a marked similarity in personality trait perceptions. This could be impacted by the average score of males in the high BPD group 19.89, far below the 45.72 average female score. This is likely to be driven by the small male sample size within the photo shoot compounded by the extreme female bias in BPD, it is predicted that 76% of BPD patients are female (Widiger & Weissman, 1991) and as a consequence observers are less likely to have encountered males with BPD, thus are less likely to have learnt cues to aid recognition.

One noticeable difference between the findings is the perception of the low BPD male image to be more attractive than the high. However, due to the halo effect (Dion, Berscheid, & Walster, 1972) it would be expected that this finding would encourage the attribution on BPD symptoms to the high BPD image Additionally, once corrected for multiple comparisons only results for Extraversion in females remained. In previous studies low Extraversion scores have been continuously associated with accurate discriminations, suggesting that it may be important in the recognition of BPD as well. If a Likert scale rating of BPD symptoms was utilised then a regression analysis could be utilised to determine if the variance in BPD symptom attribution were attributable to any other rating, such as attractiveness or Extraversion.

Although observers were unable to accurately discriminate between high and low BPD male images, they did attribute similar personality traits as with female images. These could have potential social implications for individuals with BPD, as studies have indicated that friends of highly extraverted individuals rate their friendships as closer and friends of highly agreeable individuals report less irritation (Berry et al., 2000). These social implications could be further emphasised when interacted with the symptoms of BPD. For example the DSM criteria for BPD includes compromised ability to recognise the feelings of others as well as unstable and conflicted close relationships (American Psychiatric Association, 2013).

These results suggest that observers are able to accurately determine between facial images of females scoring high and low on symptoms of borderline personality disorder. In addition high BPD images of both sexes were perceived to be lower in Agreeableness and Extraversion, supporting associations identified in the five factor model of personality disorders. It must be noted that a student sample was used in this study, and although promising results were developed they would benefit from being replicated with a clinical sample.

The next step

Here I demonstrated that another common psychiatric disorder, BPD, is observable from static facial images of females, however results were not replicated for males. Once more negative personality traits were attributed to high scorers, even when individuals were unable to accurately detect the core psychiatric trait. In the following chapter I move on to assess if a common, but neurodevelopmental disorder, autistic spectrum disorder, is observable in a similar manner.

Chapter six: Autism in the face

Abstract

Autistic spectrum disorder (ASD) is a common disorder thought to currently affect around 1% of the UK population (Brugha et al., 2011). It has been repeatedly associated with an increased prevalence of minor physical abnormalities, especially in the facial region (Ozgen et al., 2011). This study aimed to determine if these could be utilised as cues in the recognition of ASD from static facial images. When stimuli created from the highest and lowest scorers on the Autistic Quotient (Baron-Cohen, 2001) were presented to naïve observers it was found that they were able to accurately discriminate between the two groups of female stimuli, but were significantly inaccurate in their catergorisation of male stimuli. It is proposed that the inaccuracy in the ability to detect symptoms in males is due to the attribution of Extraversion to higher AQ scorers - a concept not in-tune with social perceptions of individuals with ASD.

6.1. Introduction

Autistic spectrum disorder (ASD) is a spectrum of developmental disorders which, as classified by the DSM-IV (American Psychiatric Association, 1994), causes impairment in social interaction and communication as well as restricted repetitive and stereotyped patterns of behaviour, activities and interests. ASD is thought to affect around 1% of the UK child and adult population (Baird et al., 2006; Brugha et al., 2011). Classically, it is thought to affect four times as many males as females (Bailey et al., 1995; Rodier et al., 1997), although a recently commissioned survey proposes a 9:1 male to female ratio, with 1.8% of the UK male population and only 0.2% of females being diagnosed (Brugha et al., 2011). These figures show a marked increase in ASD prevalence from the 1 in 5000 estimated in 1975 (Weintraub, 2011), although the cause of this is often disputed. It is commonly thought to be the product of an increase in awareness, however this appears to only account for 15% of the variance. Other, less well know causes include an increase in dual diagnosis with a general learning disorder (25%) and an increase in parental age (10%) (Weintraub, 2011).

ASD encompasses a range of developmental disorders, causing mild to severe impairments. An important discrimination to make regards the separation of autism and Asperger's syndrome, terms often used interchangeably. Asperger's syndrome is often portrayed as a 'mild' case of autism, both groups exhibit similar impairments in social interaction and behaviour, however individuals with Asperger's disorder do not have a delay in language or cognitive development (American Psychiatric Association, 1994). Therefore the study of ASD in non-clinical samples often uses participants scoring highly of measures of Asperger's disorder, such as the Autistic Spectrum Quotient (AQ, Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). This is a self-completed questionnaire which assesses levels of impairment in five domains;

communication, social skills, attention to detail, attention switching and imagination, with higher general scores suggesting higher levels of disability.

Autism is seen as a ‘hidden disorder’ (Autism Education Trust, 2015), contrary to other developmental disorders such as downs syndrome which have distinctive observable phenotypes. This study aims to investigate the ability to discriminate between individuals scoring highly and lowly on the AQ by viewing facial images developed from university students photographs.

There has been a series of studies conducted aiming to determine distinct facial morphology associated with ASD. Commonly using the Waldrop scale these studies have consistently shown a significant increase in abnormality scores in ASD groups when compared to controls. Rodier et al. (1997) identify the posterior rotation of ears and a shortened interpupuilary distance as setting ASD groups apart from controls. Similar results come from Walker (1997) where the ASD group had low-seated ears. A recent meta-analysis shows the consistency of increased Waldrop score in individuals with ASD, finding a pooled effect size of 0.84 (Ozgen et al., 2010). The findings of a specific facial morphology in individuals with ASD suggests a possible mechanism for symptom recognition from faces, however much of this research is based on the Waldrop scale, which only focuses on specific locations leaving questions about its real world application.

Using more modern techniques Aldridge et al. (2011) has extended this field of research. After taking 3D photos of boys aged 8-12 with and without autism statistical analysis was conducted showing significant differences in facial morphology between the two groups. From this they determined a distinctive facial phenotype associated with autism constituting of increased breadth of mouth, orbits and upper face, a flattener nasal bridge alongside reduced height of the philtrum and maxillary region (upper jaw). Aldridge et al. (2011) suggest that these features may be caused during foetal development, as the developing face uses the brain as a foundation for facial development. If there is abnormal development of the brain, as suggested by genetic studies this could have lasting impact on facial development.

This theory can be linked to others attempting to explain the biological causes behind ASD. One such theory is Baron-Cohen's masculinity hypothesis (Baron-Cohen et al., 2001). This suggests that autism is the product of an over-masculinised brain, caused by overexposure to testosterone in-utero. If true, this theory could explain the high male : female ratio in ASD. Baron-Cohen draws evidence from multiple directions for this theory, one of which is the distinction between empathising and systemising. Empathy is classically thought of as a feminine trait, whereas systemising is predominantly male. Lawson et al. (2004) shows how this gender divide is less clear-cut in individuals with ASD with female controls scoring significantly higher on empathising tasks than male controls, who in turn scored significantly higher than males with autism.

Further evidence derives from studies of testosterone related symptoms in women with ASD. One such example comes from Ingomnukul, Baron-Cohen, Wheelwright and Knickmeyer (2007) who showed a self-reported increase in testosterone related medical conditions such as irregular menstrual cycles, bi or asexuality and an increase in polycystic ovaries in women with ASD. The masculinity hypothesis can also be looked at from a bottom up perspective, looking at the prevalence on ASD in women with increased pre-natal testosterone levels. Studies following up children who had amniotic foetal testosterone samples have consistently shown an increase in ASD prevalence in children with high foetal androgen levels (Auyeung et al., 2009; Ruta et al., 2011). Finally support also comes from the study of females with congenital-adrenal hyperplasia, a genetic disorder which causes the overproduction of foetal androgens. Females with this disorder exhibit an increase in masculine symptoms alongside a significantly higher AQ score than controls (Knickemeyer et al., 2006).

There is obvious support for Baron-Cohen's male brain hypothesis, but could this form a basis for the identification of ASD from the face? Testosterone has multiple effects on the body, including an impact on facial morphology during puberty. It stimulates the widening of mandibles, forward growth of eyebrow ridges, development of larger cheekbones and lengthening of the lower facial bone (Fink & Penton-Voak, 2002). A combination of this morphology could form a possible mechanism for the identification of ASD.

A final possible influence on the identification of ASD from the face is the impact of personality. Recent research has outlined the ability to identify three of the Big Five personality traits – Neuroticism, Extraversion and Agreeableness from composite facial images (Kramer & Ward, 2010). Interestingly a specific combination of these personality traits is associated with ASD. Participants scoring highly on the AQ also reported high levels of Neuroticism and low levels of Extraversion and Conscientiousness (Wakabayashi et al., 2006). Similar findings are reported by Austin (2005), although low levels of Agreeableness replace findings of Conscientiousness here.

Using the research outlined it can be hypothesised that ASD is associated with abnormal brain development which impacts on foetal facial development, alongside the developmental effects of testosterone. Together these could form a distinct phenotype associated with ASD. A secondary theory regarding the identification of ASD in the face surrounds personality traits, personality traits are accurately identified and there is a consistent personality trait combination associated with the disorder – is it possible that these are also cues? This study aims to explore these hypotheses, determining if participants are able to accurately discriminate between facial images of individuals scoring high or low on the AQ and explore the possible reasoning for this ability.

6.2. Method

This study has two primary phases, the first of which involved the collection of photographic stimuli alongside participant AQ scores. These were then used to develop the stimuli used in the second study phase, a rating experiment.

6.2.1. Phase one: Stimulus creation

In this phase photographs of participants were collected alongside their AQ score. These photographs were then sorted and the highest and lowest 15 male and female scores were selected to be morphed together using the JPsychomorph programme (Tiddleman, Burt, & Perrett, 2001) to create composite images. Following this the composite images were used as anchor points to warp individual photos to represent the shape and texture of the high and low composites.

6.2.1.1.Participants: Photographs of 225 Bangor University students (130 females, age $M = 21.45$, $SD = 5.04$) were collected in photo shoot A (See appendix B for details) and participants were paid £5 for their participation.

6.2.1.2. Measures: The degree of ASD symptoms was assessed by the AQ (Baron-Cohen et al., 2001). This is a 50-item questionnaire with five subscales depicting different symptoms clusters; Communication, social skills, attention to detail, attention switching and imagination. Each subscale represents ten questions, with a maximum score of ten, resulting in a total possible questionnaire score of 50. Participant's scores ranged from 3-41 for males and 6 – 35 for females, with a mean of

17 and 16 respectively. Participants also completed the mini-IPIP (Donnellan et al., 2006), this 20-item questionnaire assesses an individuals personality traits.

6.2.1.3. Procedure: Participants were asked to remove all make-up and jewellery and tie back their hair before a facial photograph was taken whilst they adopted a neutral facial pose. This was taken against a light background whilst the camera distance (2m), height, flash and zoom were kept constant.

Non-white participants, those aged over 30 and males with facial hair were removed from the sample at a later date, resulting in a sample of 106 females and 48 males.

The highest (males score range 18-32, females 21-34) and lowest (males 3-13, females 6-10) 15 AQ scorers were then identified and using the Jpsychomorph programme (Tiddleman et al., 2001) morphed together to create composite images. This process resulted in four images, one representing individuals with high AQ scores and one for low scores for both males and females, as shown in Figure 6.1. Next these images were used as anchor points to warp individual images to the two extremes they represented morphologically.

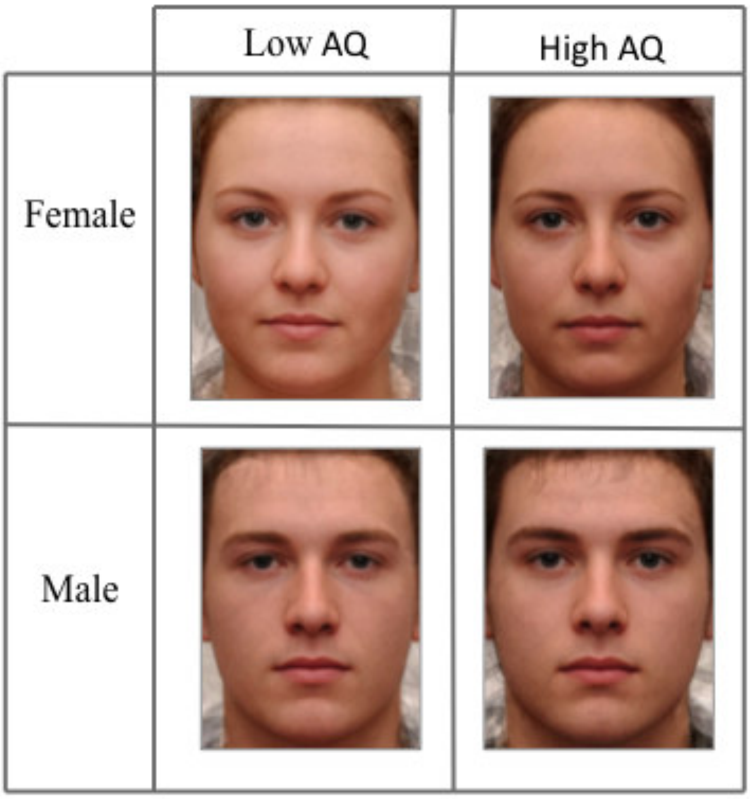


Figure 6.1. Examples of the composite stimuli created from the highest and lowest scorers of the AQ.

Twenty male and twenty female images, with average AQ scores, who had consented to having their photo used individually were randomly picked from the dataset and warped 50% towards each of the two composite images (Rowland & Perrett, 1995). This procedure resulted in two images from each photo, one representing the structural features and texture of the individuals who made up the high composite and the other the low composite images (Figure 6.2).

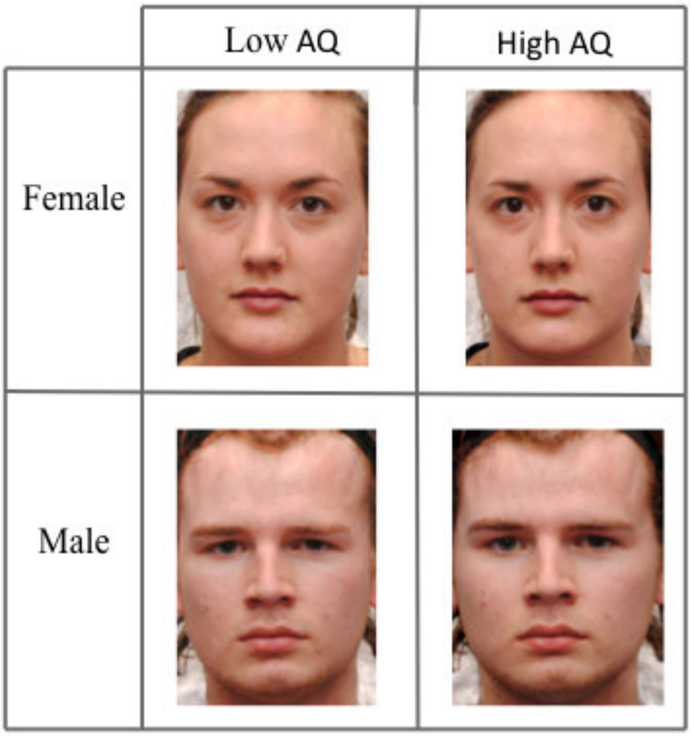


Figure 6.2. Examples of stimuli created by warping individual photos 50% towards the high and low composite stimuli illustrated in figure 6.1.

An additional set of 80 warped pairs was created using the same methodology as above, except using the highest and lowest 15 scorers from two subsections of the AQ – communication and social skills (Figure 6.3.) These 120 stimuli pairs were then used in the following rating experiment. This additional stimulus was included as these subsections most strongly correlate with deficits in interpersonal skills. As they are important in social interactions it was sought to determine if there would be a higher accuracy in comparison to general high AQ scorers.

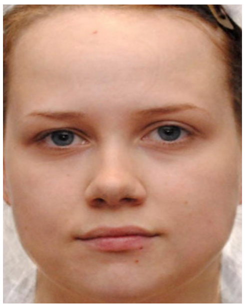
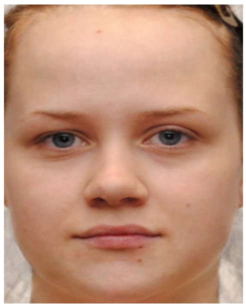
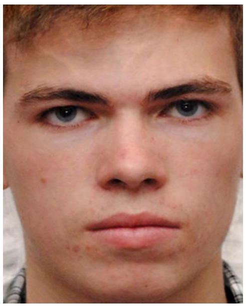
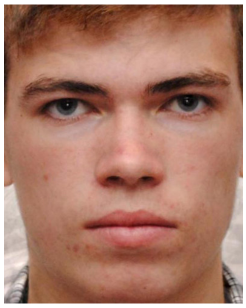
	Sex	High ASD	Low ASD
Social skills	Male		
	Female		
Communication	Male		
	Female		

Figure 6.3. Figure illustrating warped images created from the composite images of high and low ASD sub-scale scorers.

6.2.2. Phase two: Rating experiment

In this phase of the experiment the stimuli developed in phase one were presented to participants with questions designed to assess their ability to discriminate between individuals scoring high and low on ASD measures. Additional questions explored personality attributions.

6.2.2.1. Participants: Thirty-eight (23 female) participants (aged 18-38, *M* = 20.31, *SD* = 3.18 years) from Bangor University were compensated with course credits for their participation.

6.2.2.2. Procedure*:* Participants were presented with each warped pair side by side (position counterbalanced) below a description. Participants were then asked to decide which photograph best fitted the description provided and demonstrate their decision by clicking on the appropriate image with the computer mouse. Participants were presented with each of the 40 AQ warp pairs on five occasions, judging them on the following themes: ASD (Has more symptoms of autism: people with autism like routines, find it hard to maintain peer relationships, have restricted interests), attractiveness (More attractive in appearance), Neuroticism (More neurotic (depressed and blue, worries more, gets nervous easily, moody), Extraversion (More extraverted (talkative, enthusiastic, assertive, outgoing and sociable) and Agreeableness (More agreeable (co-operative, kind and considerate, generally trusting, helpful and unselfish), resulting in 200 trials. In an additional 40 trials the social skills warps were presented and participants asked 'Has poorer social skills'. A final 40 trials consisted of viewing the communication problem warps and deciding 'has more problems communicating with others'.

This resulted in a 280 trial experiment, where image pairs were presented in a random order over two blocks. There was no fixed viewing distance or time constraints on discriminations.

6.3. Results

Participants were able to accurately discriminate between the high and low AQ images in 55%, 95% CI [51-60%] of trials with female faces, $t(37) = 2.29$, $p = .028$, $r = .35$, however they were significantly inaccurate in their judgements of males, $t(37) = -2.29$, $p = .028$, $d = .39$, picking the high AQ face to have more symptoms in only 44%, 95% CI [39-49%] of trials. They also performed significantly above chance levels when judging both the male (57%, 95% CI [51-62%]), $t(37) = 2.57$, $p = .014$, $d = .42$, and female (67%, 95% CI [62 - 72%]) , $t(37) = 6.97$, $p < .0000001$,$d = 1.14$, communication problems warps. However they performed at chance levels when judging the male social skills warps, $t(37) = -0.06$, $p = .95$, $d = .48$, only picking the high AQ face in 50%, 95% CI [46 - 54%] of trials, and were significantly inaccurate when judging females, $t(37) = -2.11$, $p = .042$, $d = .34$, picking the high AQ face in 46.5%, 95% CI [43 - 50%] of the trials.

As shown in figure 6.4 participants also rated the ASD warps for personality traits. The high ASD male warps were judged as significantly more Extraverted, $t(48) = 2.58$, $p = .014$, $d = .42$, but judgements of Neuroticism and Agreeableness failed to rise above chance levels ($p > .05$). Judgements of the female warps provides a very different picture, with high ASD images being judged as significantly less Extraverted, $t(48) = -4.01$, $p = .00029$, $d = .65$, and Agreeable, $t(48) = -2.80$ $p = .008$, $d = .45$, but higher in Neuroticism, $t(48) = 6.31$, $p < .000001$, $d = 1.02$. Finally, participants also judged the ASD warps for attractiveness. High male ASD warps were judged as significantly more attractive, $t(48) = 2.35$, $p = .024$, $d = .38$. However judgements of females failed to reach significance ($p > .05$). [8]

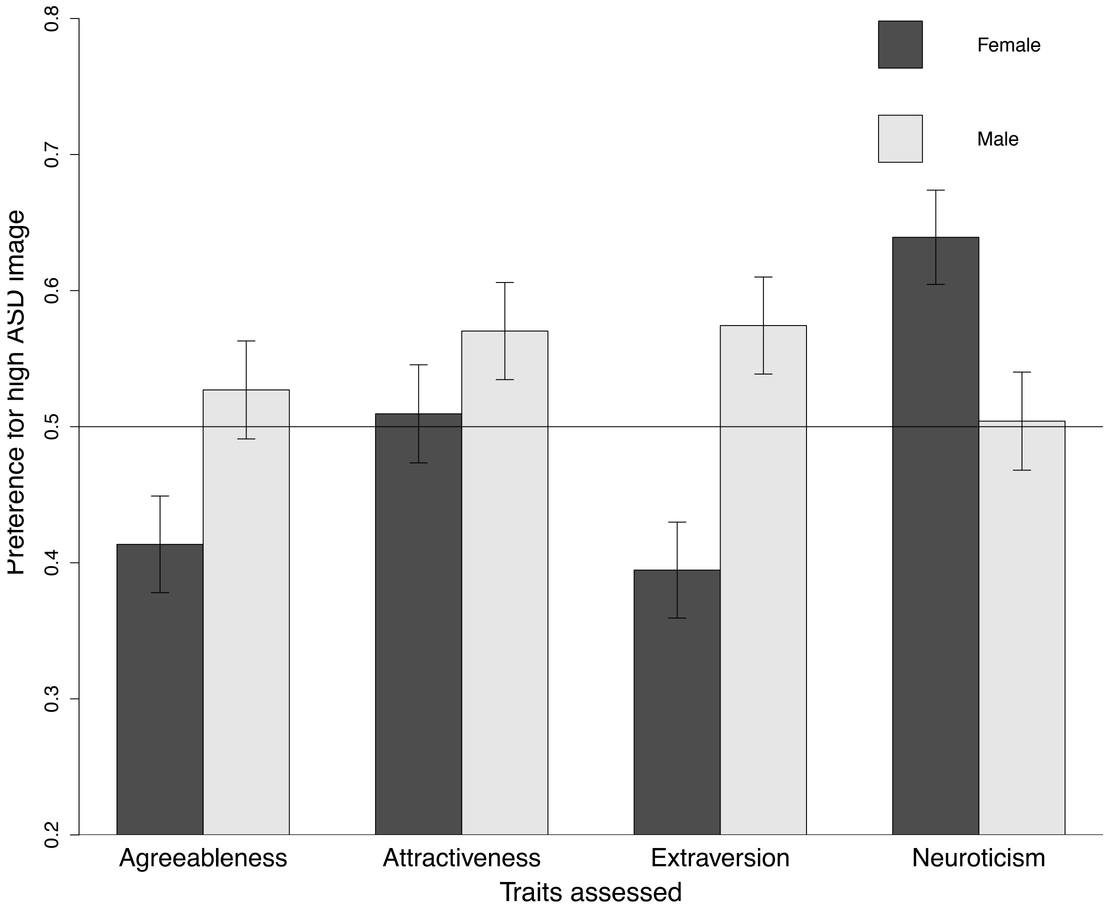


Figure 6.4. Figure detailing how often observers picked the high ASD image to be higher in each personality trait. Error bars represent 95% confidence intervals of the mean.

In order to identify the possible drive behind the discrepancy between ASD ratings it was determined if any personality attributions showed a consistent association with the ASD ratings. As shown in Table 6.5, it was found that, for both genders,

---

[8] Please note when correcting for multiple comparisons results for Attractiveness in Males fails to reach significance ($p = .096$) and results for Extraversion reach only marginal significance ($p = .07$).

ratings of Extraversion were significantly negatively correlated with ASD ratings. This suggests that observers who picked the high ASD face as higher in Extraversion also picked it to be lower in symptoms of ASD. When the correlation between ratings of Extraversion and ASD are taken into consideration it suggests that cues to Extraversion may be important in the perception of ASD and may help explain why observers were inaccurate in their perceptions of ASD in males.

Table 6.5. Table detailing the correlations between ratings of ASD and personality traits.

	Attractiveness	Agreeableness	Extraversion	Neuroticism
Female	-.08*	-.02	-.11*	.03
Male	-.04	-.07	.13*	.07

6.4. Discussion

This study demonstrated that naive observers were able to accurately determine between facial images of females who scored high or low for ASD traits but were significantly inaccurate in their discrimination of male faces. Correlation analysis assessing how often observers picked the high ASD face to be higher in desirable personality traits suggested that Extraversion may be an important cue or correlate of symptoms of ASD.

Discriminations of Extraversion and ASD were found to be negatively correlated in female trials but positively correlated in trials with male stimuli. That is, observers would consistently rate the image they perceived to be low in Extraversion as higher in ASD in females, where they were also highly accurate in their discrimination of ASD symptoms. However in trials with male stimuli observers would pick the image they perceived to be high in Extraversion as having more symptoms of ASD, but were inaccurate in their discrimination of ASD. Thus, they perceived the high ASD image to be higher in Extraversion but lower in symptoms of ASD than the low ASD image.

This opens the question of why participants seem to be consistently inaccurate in their perceptions of Extraversion in males. Participants in phase one completed the mini-IPIP alongside the AQ. Taking into the accounts of Austin (2005) the 15 participants used to create the high AQ composite images should also report low levels of Extraversion. However, non-significant correlations between AQ and Extraversion were found for the males used to create both the high and low composite images. This was not the case for females, where the AQ scores from participants used to create the high AQ composite image were significantly correlated with low levels of Extraversion. A marginally significant finding was reported with the low AQ participants. Also interesting to note is that when all male participants were included there was a significant negative correlation between AQ and Extraversion scores.

These findings could suggest why participants were inaccurate with judgements of males. Whilst they expected the high AQ individuals to also exhibit low levels of Extraversion, this was not the case. Whilst this may suggest a reason for participants to perform at chance levels when judging these images – as they may be randomly guessing due to loosing their cue. However it does not explain why they are significantly inaccurate in their discriminations.

The answer could be found in the other judgements made. The high ASD male images were consistently rated as more attractive. These judgements could be forming a halo effect, where attractive individuals are attributed positive traits (Dion, Berscheid & Walster, 1972). The perception of the high ASD images as being more attractive could influence the other variables, such as Extraversion – traits seen as highly desirable. This is reflected in the correlation analysis as judgements of attraction form the second most common association, although this failed to reach significance. Future research should assess observers perceptions of these traits with Likert style rating scales, this would allow for regression analysis determining which traits, if any, are used by observers to make inferences about ASD traits, an analysis not possible with the methodology utilised here.

The perception of female high ASD faces as being more Neurotic, but less Extraverted and Agreeable than low ASD faces mimics the personality profile suggested by Austin (2005). The addition of the finding that high ASD male faces were misattributed with low ASD symptoms but seen as higher in Extraversion suggests that personality traits may be a key cue in the perceptions of ASD from facial features. If minor physical anomalies (MPAs) were to be uniquely utilised as cues then we could expect a more unified ability across the sexes. However further research, such as measuring MPAs and gaining judgments of ASD from faces with differing Waldrop scores would be required to make a causational inference about the application of these cues.

Although it is hard to tease apart the exact reasoning for these judgements, the findings that personality perceptions are a key driver in the identification of ASD in the face casts a new light on the social perception of ASD. It would appear that individuals have a preconception of what they expect an individual with ASD to be like. This could have a detrimental effect on individuals with ASD socially, as the personality type combinations perceived are ones seen as less desirable in friendships. For example friends of highly extroverted individuals report a closer relationship and friends of highly agreeable individuals report less irritation (Berry et al., 2000). This perceived personality make-up could produce a social disadvantage for individuals with ASD, whilst they find it hard to understand the intricate details of social interaction they are also less likely to be approached due to these assumptions resulting in a negative feedback cycle.

Participants were also presented with images representing high and low levels of two subscales of the AQ – Social skills and communication. Whilst there was a failure to find any accuracy in the judgements of the social skills scale there was a high level of accuracy in judgements of communication problems. The social skills subscale includes questions such as ‘I would rather go to the library than a party’ and ‘I enjoy social occasions’ which could be seen as questions that may not accurately identify ASD in university students, whereas the communication subscale highlights traits which may be harder to learn how to overcome, a technique adopted by many with ASD to cope with social situations. These questions include ‘Other people frequently tell me that what I’ve said is impolite, even though I think it is polite’ and ‘I frequently find that I don’t know how to keep a conversation going’.

The next step

In this chapter I demonstrated that observers were able to accurately detect ASD in female faces, and the personality attributions we have seen throughout the previous chapters was once again replicated. The results for males however was less straightforward, with observers misattributing traits to low ASD scorers, which I suggest that this result is caused by the misattribution of Extraversion to high scorers. In the next chapter I test if these findings can be explained using an existing theory of ASD, the extreme male brain theory.

Chapter seven: Facial dimorphism in Autistic Quotient Scores[9]

Abstract

Baron-Cohen's extreme male brain theory proposes that autism results from elevated prenatal testosterone levels. Here we assess possible correlated effects of androgen exposure on adult morphology, and in particular, the development of facial features associated with masculinity. We created composite images capturing statistical regularities in facial appearance associated with high- and low-AQ scores. In three experiments, we assessed correlations between perceived facial masculinity and autistic-spectrum quotient (AQ) scores. In Experiment 1, observers selected the high-AQ males as more masculine. We replicated this result in Experiment 2, using different photographs, composite image methods, and observers. There was no association of masculinity and AQ scores for female's faces in either study. In Experiment 3, we created high and low-AQ male composites from the five AQ subscales. High-AQ images were rated more masculine on each of the subscales. We discuss these findings with respect to the organisational-activational hypothesis of testosterone activity during development.

---

[9] This chapter appears in print:
Scott, N. J., Jones. A. L., Kramer, R. S. S. & Ward, R. (2015). Facial dimorphism in Autistic Quotient scorers. *Clinical Psychological Science, 3*(2), 230-241.

7.1. Introduction

The Extreme Male Brain theory of autistic spectrum disorder (ASD; Baron-Cohen, 2002) assumes a continuum of individual differences in social and cognitive style, ranging from an empathising "female" brain to a systematising "male" brain. The empathising end of the spectrum is represented by social cognitive skills, such as identifying the emotions and thoughts of others. The systematising end of the scale is represented by skills in analysis of the underlying constructs in mechanical (i.e., non-social) systems. Baron-Cohen identifies many behavioural traits that support the gender assignments of this continuum; for example, females are more likely to exhibit empathic behaviours such as responding to the distress of others, taking turns, and showing sensitivity to the facial expressions of others. In turn, males are more likely to have good mathematical and engineering skills; pay more attention to relevant details of mechanical systems; and show a greater aptitude for construction (Baron-Cohen, 2002). Therefore, within the context of a broad range of individual differences, Baron-Cohen argues that the average woman will lie more towards the empathising end of the scale, and the average man more towards the systematising end.

Individuals with ASD have been shown to exhibit impaired empathising abilities and high degrees of systematic behaviour in various ways. Lawson, Baron-Cohen and Wheelwright (2004) found that, as expected, females outperformed males on empathising tasks; however, males with Asperger disorder scored even lower than control males. Similarly, both males and females with ASD were found to have lower scores on the Empathising Quotient than controls (Stauder, Cornet, & Ponds, 2011). The Empathising Quotient asks participants to respond to questions such as ‘I really enjoy caring for other people’ on a four point scale from ‘strongly agree’ to ‘strongly disagree’. Participants are given one or two points for each answer that reflects high empathising skills, resulting in a possible score of 80, where high scores represent high empathising skills (Baron-Cohen & Wheelwright, 2004). In contrast, Baron-Cohen (2002) compiled further evidence for improved systemising skills with increasing ASD severity, such as a preference for information which is rule-based, structured and factual; being more likely to undertake science rather than arts degrees; and for males, higher scores on the Systemising Quotient for ASD groups compared to controls (Baron-Cohen, 2002). The Systemising Quotient is a measure with similar format to the Empathising Quotient, where participants are instead asked questions such as 'I prefer to read non-fiction than fiction' (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003).

Extreme Male Brain theory: Prenatal testosterone and ASD. The Extreme Male Brain theory proposes that the highly systemising behaviour seen in ASD results from over-exposure to androgens during prenatal development, which in turn produce a highly masculinised brain. This account is consistent with findings showing that high foetal androgen levels, based on amniotic fluid samples, are associated with increased ASD occurrence (Auyeung et al., 2009; Ruta, Ingudomnukul, Taylor, Chakrabarti, & Baron-Cohen, 2011). The androgen over-exposure postulated by Extreme Male Brain theory predicts not only differences in socio-cognitive style, but also other effects downstream of androgen exposure. One example would be the relatively high male:female ratio in ASD, where diagnoses are three to five times more common in males than females (Bailey et al., 1995; Cosgrove & Riddle, 2004; Fombonne, 1999; Rodier, Bryson, & Welch, 1997). In a recently commissioned UK survey, where adults were screened using a shortened version of the Autistic spectrum quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) this ratio increased to nine to one (Brugha et al., 2011). Some other effects linking testosterone-related symptoms with the occurrence of ASD have been documented. Ingudomnukul, Baron-Cohen, Wheelwright and Knickmeyer (2007) found that females with ASD were more likely to report testosterone-related medical conditions, such as irregular menstrual cycles and an increased risk for polycystic ovary syndrome, additionally they were more likely to be bi- or asexual. Furthermore, females with congenital-adrenal hyperplasia, a genetic disorder which causes the overproduction of foetal androgens, exhibit an increase in masculine traits assessed by the AQ, alongside a significantly higher overall score compared to controls (Knickmeyer et al., 2006).

More generally, androgen exposure through the lifespan is associated with a variety of morphological, as well as cognitive developments. In the present study, we explore whether facial masculinity is a visible phenotype correlated with autistic traits. We will first examine whether there is a case that the prenatal androgen exposure, assumed by Extreme Male Brain theory, may be correlated with effects on facial masculinity. We then present three studies which test this correlation.

Facial masculinity and levels of prenatal and pubertal androgens. It is well known that pubertal androgen levels affect facial masculinity. Facial masculinity in boys normally begins to develop at puberty, along with other rapid changes in body shape associated with sexual dimorphism, such as increases in body size, and this adolescent development is associated with a large increase in androgen levels of boys (Tanner, 1989). The causal role of testosterone in facial masculinity is suggested by the fact late developing boys given testosterone treatment develop longer total mandibular length and anterior face height (Verdonck, Gaethofs, Carels, & de Zegher, 1999). Furthermore, facial shape features used by observers to identify masculinity in pubertal adolescents include features correlated with circulating testosterone levels in male adolescents, such as increased breadth of the forehead, jaw, and chin (Mareckova et al., 2011).

However, the Extreme Male Brain theory predicts that it is effects of prenatal, not pubertal, androgens on brain development that promote ASD. What is less clear is whether there is an effect of prenatal androgens on facial masculinity. It is important to be clear that androgen levels are not stable during development, but are better described by a two-stage process. According to the long-standing organisational-activational model (Phoenix, Goy, Gerall, & Young, 1959), sex hormones first establish primary sex characteristics and the “organisation” of sexually dimorphic structures, including aspects of the brain, during perinatal development. Outside primary sex characteristics, this organisation may not be readily apparent until “activation” during the second stage, occurring at puberty (see Wallen, 2009, for a review of developments). It is therefore plausible that facial masculinity, like other sexual dimorphisms such as body size (Chowen, Argente, Gonzalez-Parra, & Garcia-Segura, 1993), might be influenced by prenatal androgen levels during this organisation period, even if the effect is dormant until activation during puberty.

We tested whether facial masculinity is associated with ASD, that is, whether facial masculinity is a downstream consequence of, or correlate to, prenatal androgen exposure. A positive correlation between facial masculinity and ASD would support a novel prediction of Extreme Male Brain theory, linking behaviour and morphology through linked hormonal effects. In addition, a positive correlation would also illuminate what are currently poorly understood pathways for the effects of prenatal androgens on adult sexual dimorphism. Regardless of the mechanisms involved, identifying a visually salient phenotype relating to ASD, like facial masculinity, may be important for better understanding the social perceptions that are linked to ASD.

Facial morphology and ASD. Previous research has investigated the concept of facial traits associated with ASD. These studies have typically used the Waldrop scale, which lists specific abnormalities and corresponding weighted scores, to gain a standardised score for minor physical abnormalities (MPAs; Waldrop, Pedersen & Bell, 1968). For example, Rodier et al. (1997) and Walker (1977) found a significant increase in MPA scores for children with ASD compared to controls. In a comprehensive meta-analysis Ozgen, Hop, Hox, Beemer and van Engeland (2010) demonstrated a significant association between MPAs and ASD with a pooled effect size of $d = 0.84$. Ozgen et al. (2011) and Ozgen et al. (2013) confirmed this association between ASD and MPA in subsequent, more controlled studies. These studies do not argue that a specific anomaly or set of anomalies is diagnostic of ASD, but that on aggregate, a higher MPA score is more likely to be associated with ASD.

Other work has investigated specific facial features that may be diagnostic of autistic traits. Aldridge et al. (2011) found significant differences in the facial morphology of boys aged 8-12 years with and without autism, including increased breadth of mouth, orbits and upper face, a flatter nasal bridge, and reduced height of the philtrum and maxillary region. Aldridge et al. suggest that as the developing face uses the brain as a foundation for craniofacial development, influences on brain development could produce corresponding, lasting impact on facial appearance. Hammond et al. (2008) found greater facial asymmetry in boys with ASD compared to controls, and also the ability to discriminate between the faces of boys with ASD and controls using pattern-matching algorithms. These findings support the notion of a specific phenotype associated with ASD, but do not link this type to perceptions of masculinity. Finally, Bejerot et al. (2012) asked eight observers to assess the face and body of ASD and control individuals, using a 5-point coherence-typicality scale which coded for “gender typicality” at one end, and “gender coherence” at the other, reported by Bejerot et al. as: very gender typical; gender typical; average; weak gender coherence; very weak gender coherence. ASD female faces were rated as significantly less gender coherent than controls, but no significant differences were found for male faces. Unfortunately for our purposes, this scale entangles the typicality of traits with their dimorphism. For example, an average male face would, by statistical definition, be most typical of males (i.e., an extreme score of typicality); while a hyper-masculine face would, again by statistical definition, be less typical of males (i.e., a less extreme typicality score). This coherence-typicality scale therefore does not seem to directly address the question of whether there is an association between autistic traits and facial dimorphism.

Here we test the specific hypothesis of association between facial masculinity and autistic traits. As a non-clinical population was tested, autistic traits were measured using the AQ (Baron-Cohen et al., 2001). We discuss the AQ measure in more detail below. We created two databases, each consisting of over 200 participants facial photographs and AQ scores. From each database we selected the males and females with the highest and lowest AQ scores, and made composite, or “average” images, e.g., the high-AQ male, the low-AQ female. These composites captured statistical regularities in facial appearance associated with high and low AQ, while filtering out facial idiosyncrasies. We then used the composites to create stimuli varying on facial properties associated with AQ score, and measured observer responses for masculinity.

Our hypothesis is that correlates of prenatal testosterone may affect the masculinity of the adult face in men with high AQ scores. This prediction follows from studies examining the morphological consequences of testosterone exposure, conducted using males (e.g., Schaefer, Fink, Mitteroecker, Neave & Bookstein 2005; Penton-Voak

& Chen, 2004; Sisk, Schultz & Zehr 2003; Verdonck et al., 1999). For example, adolescent testosterone is known to affect masculinity of male faces (Mareckova et al., 2011). Predictions for female faces are much less clear, because at present it is not known how testosterone might affect masculinity of the female face. We have therefore included female faces in Experiments 1 and 2 as an exploratory comparison.

7.2. Experiment 1

This experiment consisted of two phases. In Phase One, we collected a dataset of participant photographs alongside their AQ scores, used to develop our face stimuli. In Phase Two, these stimuli were presented to observers to investigate perceptions of masculinity.

7.2.1. Phase One: Stimulus Creation

During this phase, we aimed to create male and female composite images which reflected the commonalities in face shape from those scoring in the extremes of the AQ. These were then used to create "high-AQ" and "low-AQ" versions of individual faces.

7.2.1.1. Participants: Two hundred and twenty-five Bangor University students (130 females, age $M = 21.45$, $SD = 5.04$) took part in photo shoot A (See appendix B for details) and were paid £5 for their participation.

7.2.1.2. Measures: The level of autistic traits was assessed by the AQ (Baron-Cohen et al., 2001). The AQ is not a diagnostic questionnaire, but used to identify the extent to which adults of normal intelligence show autistic traits and has been demonstrated to be a valid and reliable measure of autistic traits in an adult population. It is formed from 50 items, which are split into five subscales depicting different symptoms clusters; communication, social skills, attention to detail, attention switching and imagination. Each subscale is represented by ten questions which the participant answers on a four point scale, from definitely agree to definitely disagree, a point is given when the extreme answer (definitely agree/disagree) correlated with autistic traits is selected, resulting in a total possible questionnaire score of 50. A cut-off score of 32 is proposed as a clinical threshold (Baron-Cohen et al., 2001). Participants' scores ranged from 3-41 for males (*M* = 17.31, *SD* = 6.56) and 6–35 for females (*M* = 16.21, *SD* = 5.62).

7.2.1.3. Procedure: Participants were asked to remove all make-up and jewellery, tie back their hair and adopt a neutral expression whilst a photograph was taken. This was captured against a light background whilst the camera distance (two metres), height, flash and zoom were kept constant.

To control for racial characteristics and age in composite images, photographs used in the study were restricted to participants who self-reported as White or White British ethnicity, and those under 30 years old. Men with beards and moustaches were also removed, resulting in a sample of 106 females and 48 males.

The 15 highest (female scores *M* = 25.00, *SD* = 4.49, male scores *M* = 21.47, *SD* = 4.16), and the 15 lowest (female scores *M* = 8.33, *SD* = 1.35, male scores *M* = 9.40, *SD* = 3.78) AQ scorers for each sex were then identified. Using the JPsychoMorph software (Tiddeman, Burt & Perrett, 2001), the photographs of the individuals in these four groups were averaged to create composite images. This process resulted in four composite images: males with high-AQ scores, males with low-AQ, females with high-AQ, and females with low-AQ.

Next, these composites were used as anchor points to warp individual images to the two extremes they represented. Images from 20 males and 20 females, with average AQ scores (all within 1 SD of the mean), and who had consented to having their photo used individually, were chosen at random from the dataset. Each of these target faces was then warped 50% towards each of the two composite images, resulting in 40 stimulus pairs: two images for each target face, one representing the structural and surface features of the individuals making up the high-AQ composite and the other the low-AQ composite images, examples given in Figure 7.1. To create the warped images a template of facial landmarks are used as anchor points to allow the individual images to adopt the structural morphology of the composite images and the relevant colour and texture information from each facial feature is also attributed to the new image (for full description see Rowland & Perrett, 1995).

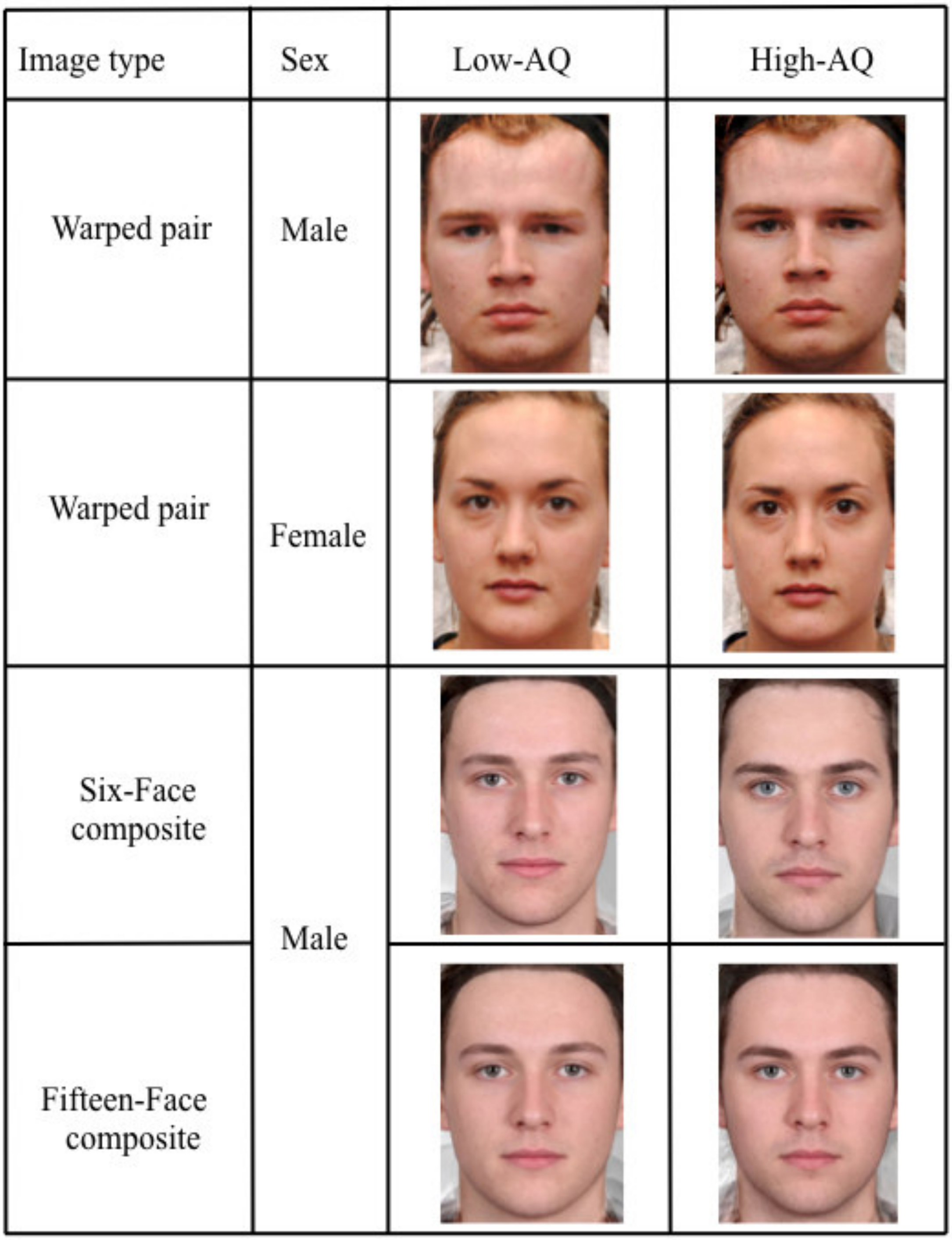


Figure 7.1. Example stimuli created from high and low scorers of the AQ, utilising a range of methodologies including warped images and composite images created from 6 and 15 individuals.

7.2.2. Phase Two: Perceived masculinity

In this phase, the images representing the high- and low-AQ versions of each target face were presented side by side, and observers were asked to pick the image they perceived to be more masculine.

7.2.2.1. Observers: Thirty-eight (23 female) observers (age $M = 20.31$ years, $SD = 3.18$) from Bangor University were compensated with course credits for their participation.

7.2.2.2. Procedure: On each trial, observers were presented with a stimulus pair consisting of the same target face warped towards the high-AQ and low-AQ composite for its sex (position counterbalanced across stimulus pairs). Observers were asked to indicate which image they thought was 'More masculine in appearance' by clicking on the appropriate image with the mouse. This then initiated the next trial.

Stimulus pairs were presented in a different random order for each participant. Responses were not speeded and viewing distance was not set. Observers were unaware of the hypotheses being tested or the criteria used for making the stimuli.

7.3. Results The frequency with which observers chose the high-AQ face to be more masculine was recorded. No difference was found between male and female observers, $F(1, 36) = 0.86$, $p = .36$, so results presented are aggregated over observer sex. As shown in figure 7.2, high-AQ male faces were selected as more masculine 69%, 95% CI [62-75%], of the time, a rate significantly higher than expected by chance, $t(37) = 6.07$, $p < .0001$, $d = .1.01$. However, observers showed no bias for the high-AQ female faces, which was selected as more masculine than the low-AQ female faces 47%, 95% CI [41-53%], of the time, a rate no different from chance, $t(37) = 0.90$, $p = .375$, $d = .16$.

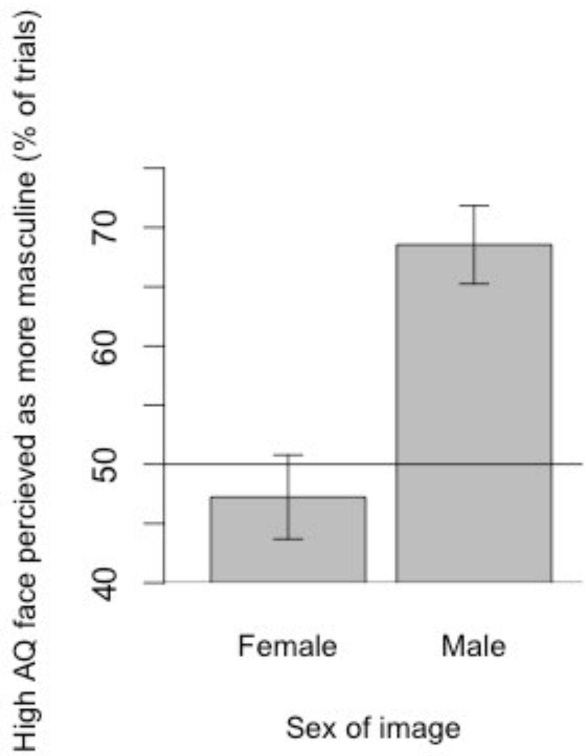


Figure 7.2. Figure detailing how observers picked the male high AQ image to be more masculine significantly more often than expected by chance, but responses to female images were at chance levels. Error bars represent 95% confidence intervals of the mean.

7.4. Experiment 2

Experiment 1 found that high-AQ men had a more masculine facial appearance than low-AQ men. We wanted to assess the validity and reliability of this finding. Our primary concern was whether unrepresentative faces in either AQ group might be driving our effects. We therefore set about testing the reliability of our results in a second experiment. Experiment 2 largely replicated the methods of Experiment 1, but with a few differences. We began by replicating the entire procedure anew; we developed a new stimulus set made from a second database of images and AQ-scores, created from a different group of participants, and then assessed by a different set of observers. Second, we changed the procedure for making the stimulus pairs. We wanted to ensure that our previous findings were not idiosyncratic to the single pair of anchors used for each sex. In Experiment 1, the high and low images (for a given sex) were warped in accordance with the differences between the single anchor pair. In Experiment 2, we therefore increased the number of comparisons, by dividing our high and low scorers into smaller subsets, creating three composites for each group. We then presented observers with all combinations of the high and the low composites.

7.4.1. Phase One: Stimuli development

The procedures of Experiment 1 were largely replicated except as noted below. Changes were mainly in the techniques used to create the new stimuli.

7.4.1.1. Participants: A new set of two hundred and twenty-one (130 females) Bangor University students (age M = 21.65, SD = 5.09) took part in photo shoot B (See appendix B for details) and were paid £6 for their participation.

7.4.1.2. Procedure: Photographs and AQ scores were collected as before. Participants' AQ scores were comparable to those in Experiment 1 (female range: 3-41, M = 17.30, SD = 6.72; male range: 3-30, M = 17.74, SD = 5.76). Again, photographs from the 15 highest (female score M = 28.20, SD = 5.19; male score M = 24.87, SD = 2.45) and lowest (female score M = 7.53, SD = 2.10, male score M = 10.20, SD = 2.60) scorers from each sex were used to create composite "fifteen-face" images, as in Experiment 1.

To ensure that results were not driven by the appearance of a few individuals used in the composite images, additional composites were created. The 18 highest (overall female score M = 27.06, SD = 5.40; male score M = 24.17, SD = 2.75) and lowest scores (overall female score M = 7.94, SD = 2.13; male score M = 10.50, SD = 2.62) for each sex were selected. These groups of 18 were randomly divided into three subsets of six faces each, and a “six-face” composite made from each subset. An individual appeared in only one six-face composite, and all individuals in the fifteen-face composite appeared in one of the six-face composites.

In total, there were 16 composite images. For each sex, there was one high and one low fifteen-face composite, and three high and three low six-face composites (examples shown in Figure 7.1). These images were then assessed for masculinity in Phase Two.

7.4.2. Phase Two: Observer impressions of masculinity

Selection of masculine faces was assessed similarly to Experiment 1, with differences noted below.

7.4.2.1. Observers: A new set of 48 (36 female) Bangor University students (age $M = 20.54$, $SD = 3.05$) were compensated with course credits for their participation.

7.4.2.2. Procedure: The main procedure of Experiment 1 was repeated with each trial consisting of a high- and low-AQ image being presented side by side. Each possible combination of the three high and three low-AQ six-face composites was presented once, for a total of nine trials using men’s faces, and nine trials using female’s. In addition, the high- and low-AQ versions of the fifteen-face composites were presented, once with the men’s faces, and once with the female’s. This resulted in 20 trials overall.

7.5. Results

The frequency with which observers picked the high AQ face as more masculine was recorded. Again, no differences in choices made by female and male observers were found, $F(1,46) = 1.66$, $p = .67$, so results were aggregated over observer sex. The results were comparable to Experiment 1. As shown in figure 7.3 over all trials, observers picked the high-AQ male face as more masculine 69%, 95% CI [63-75%] of the time, a level significantly higher than chance, $t(47) = 6.61$, $p < .00001$, $d = .95$. With the female faces there was again no significant difference, as the high-AQ image was selected as more masculine 48%, 95% CI [41-54%], of the time, $t(47) = 0.79$, $p = .43$, $d = .009$. Results from the six and fifteen-face composites were comparable. With the male composites, observers picked the high-AQ face as more masculine 77%, 95% CI [62-74%], of the time when viewing six-face composites, and 68%, 95% CI [65-89%], of the time when viewing the fifteen-face composite, both results significantly different from chance, $t(47) = 6.00$, $p < .00001$, $d = 1.28$, , and $t(47) = 4.42$, $p = .000058$, $d = 0.42$, respectively. For the six-face female composites the high-AQ face was picked as more masculine 47%, 95% CI [40 - 53%], of the time, a rate not different from chance, $t(47) = 1.03$, $p = .31$, $d = .14$; and 54%, 95% CI [40 - 69%], of the time when viewing the fifteen-face composite, again not different from chance, $t(47) = 0.57$, $p = .57$, $d = .08$. Additionally, results were generally consistent within the combination trials of six-face composites, with the high-AQ male image being picked as more masculine between 66 and 73% of the time, and the high-AQ female image being picked between 47 and 56% of the time.

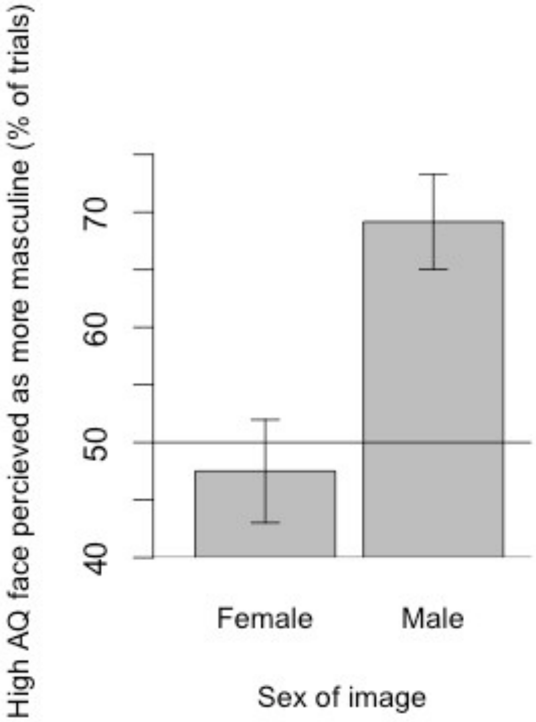


Figure 7.3. Figure illustrating the replication of the results from experiment one. Error bars represent 95% confidence intervals of the mean.

7.6. Experiment 3

In Experiments 1 and 2 we demonstrated a consistent association between high AQ scores and perceived facial masculinity in males. The AQ is composed of five subscales, relating to: social skills; communication skills; attention switching; attention to detail; and imagination. Here we looked in more detail at different components of the AQ score. In particular, is the association we find between AQ and facial masculinity in men identifiable in all of the AQ subscales? Or is the general association with AQ driven by a specific correlation of masculinity with one of the subscales? We therefore created composite images from the high and low scorers for each of the five AQ subscales. As observations of female images had been at chance levels for previous studies, only male images were used in this experiment.

7.6.1. Phase One: Stimuli development

Stimuli were developed from the same photo database as Experiment 2, and changes to stimuli development are noted below.

7.6.1.1. Procedure: Photographs from the 18 highest and lowest scorers for each of the five subscales were selected; Social skills (high-AQ $M = 4.50$, $SD = 1.86$, low-AQ $M = 0.28$, $SD = 0.46$), attention switching (high-AQ $M = 7.11$, $SD = 1.13$, low-AQ $M = 2.39$, $SD = 0.61$), attention to detail (high-AQ $M = 8.27$, $SD = 1.00$, low-AQ $M = 2.78$, $SD = 1.11$), communication skills (high-AQ $M = 4.78$, $SD = 1.11$, low-AQ $M = 0.78$, $SD = 0.73$) and imagination (high-AQ $M = 3.94$, $SD = 1.26$, low-AQ $M = 0.72$, $SD = 0.46$). As in Experiment 2 these groups of 18 were then randomly divided into sub-groups of six and 'six-face' composites were made from each subset. This resulted in 30 composite images, examples of which can be found in Figure 7.4.

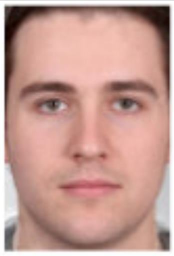
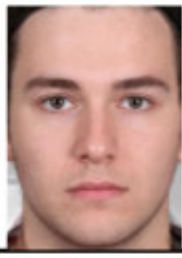
Subscale	Low AQ	High AQ
Attention to detail		
Attention switching		
Communication		
Imagination		
Social skills		

Figure 7.4. Figure showing examples of high and low AQ scoring stimuli for each of the five sub-scales.

7.6.2. Phase Two: Observer impressions of masculinity

The procedure of Experiment 2 was largely replicated, with any differences noted below.

7.6.2.1. Observers: A new set of 40 (30 female) observers was collected through opportunistic sampling (age $M = 24.70$, $SD = 4.27$).

7.6.2.2. Procedure: As in Experiment 2, each possible combination of the three high and low-AQ images for each subscale were presented side by side to observers. This resulted in nine trials per subscale and 45 trials in total. Once again, observers were asked to select the image they perceived to be 'More masculine in appearance'.

7.7. Results

There was no differences in judgements given by male or female observers for any of the subscales (all t's < .08, all p's > .37, all r's < 0.14), thus results were collapsed over observer sex. The frequency with which observers picked the high-AQ face as more masculine was recorded, and this frequency compared with a 50% chance rate. As shown in figure 7.5, Observers consistently picked the high-AQ as more masculine in appearance across all subscales: in 59%, 95% CI [54 - 64%], of trials with the attention to detail subscale, $t(39) = 3.52$, $p = .001$, $d = 0.56$, 61%, 95% CI [56 - 65%], with attention switching, $t(39) = 4.43$, $p = .0002$, $d = 0.73$, 63%, 95% CI [56-70%], with communication problems, $t(39) = 3.86$, $p =.0008$, $d = 0.61$, 83%, 95% CI [78 -87%], with imagination difficulties, $t(39) = 14.14$, $p < .00001$, $d = 2.25$, and 71%, 95% CI [64-78%], with social skills problems, $t(39) = 6.32$, $p < .00001$, $d = 1.01$(results corrected for multiple comparisons).

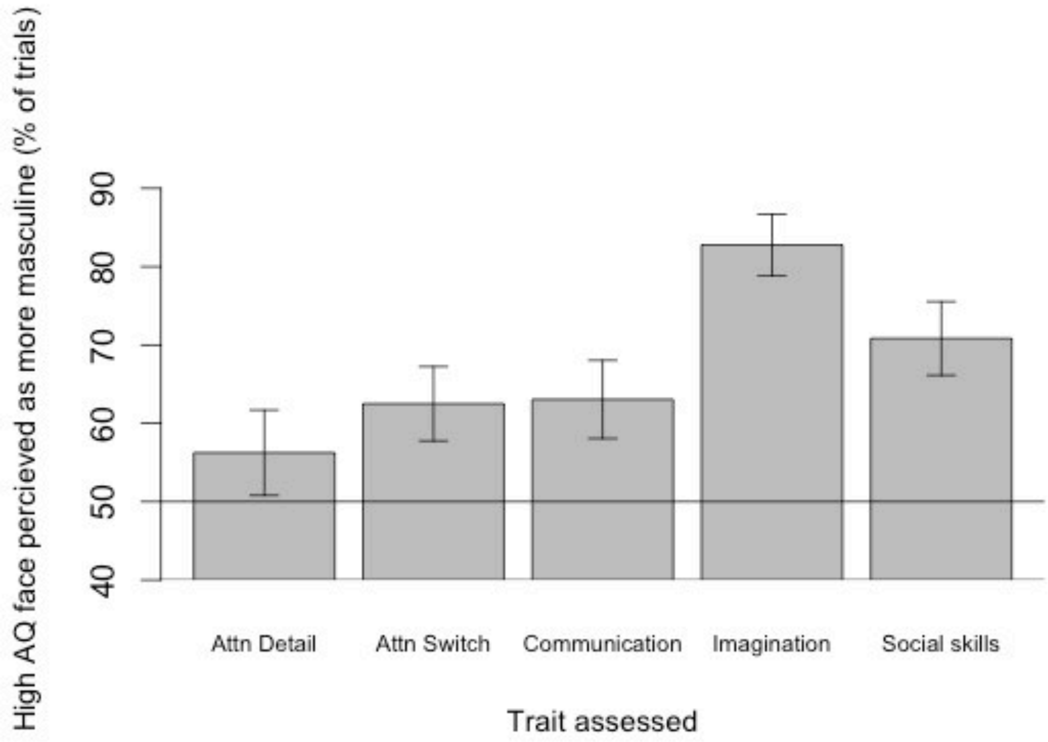


Figure 7.5. Figure detailing how the high AQ composite image was picked to be more masculine across each of the AQ sub-scales, attention to detail (Attn Detail), attention switching (Attn Switch), communication skills (Communication), Imagination and social skills. Error bars represent 95% confidence intervals of the mean.

Post-hoc analysis showed that observers rated the high AQ image as more masculine more frequently in the imagination subscale than all others (all t's > 4.98, all p's < .00001, all r's > 0.62). Secondly the high AQ social skills images were rated as more masculine with a higher frequency than the attention switching, $t(39) = 3.32$, $p = .01$, $d = 0.51$, 95% CI [0.04 - 0.17], and attention to detail, $t(39) = 3.34$, $p = .009$, $d = 0.53$, 95% CI [0.05 - 0.19], subscales (results corrected for multiple comparisons[10]).

7.8. Discussion

Our results are consistent with a shared hormonal influence on behaviour and on appearance. In our first two experiments, we found an association between facial

[10] Accuracy comparison for communication skills was marginally significant before multiple comparisons were conducted ($p = .055$).

masculinity and total AQ scores of men, such that high AQ was associated with more masculine faces. Using different datasets, participants, and stimulus creation methods, we found very similar results across both experiments. No association was found in female faces. In a third experiment we found that the association of facial masculinity and AQ score in men held across all five AQ subscales. In other words, facial masculinity was correlated with (or a consequence of) the cause of high scores for each subscale. The simplest account of this pattern is to hypotheses a shared, sex-related, basis for scores on each subscale. Post-hoc analysis showed that the high AQ image was rated as more masculine more often in the imagination subscale than any other subscale, suggesting a more robust association with masculinity, although all subscales reached significant levels. Consistent with extreme male brain theory, androgen exposure may be the underlying shared basis for these effects and the association of facial masculinity and AQ scores.

These experiments were motivated by findings and hypotheses concerning the role of androgens on ASD and on facial masculinity, at different points in development. There is a step in the causal chain that is needed to link prenatal androgen levels, hypothesised by Extreme Male Brain theory to be a driver for ASD, to pubertal androgen, which is a known driver for facial masculinity. Our preferred account at present is based on the organisational-activational hypothesis (Phoenix et al., 1959), and decades of animal work showing that prenatal androgens can affect sensitivity to activational hormones at subsequent stages of development (e.g., Wallen, 2009). Post natal sexual dimorphism, both in behaviour (e.g., Phoenix et al., 1959; Grady, Phoenix & Young, 1965) and morphology (e.g., body size, Chowen et al., 1993), is dependent upon perinatal androgens in nonhuman mammals, including primates (Thornton, Zehr & Loose, 2009). By this account, facial masculinity in men would be a joint result of both early organisational effects of prenatal androgens and later activational effects of pubertal hormones.

The robust association we found for autistic traits and masculinity of men's faces did not hold for women's faces. In fact, if organisational effects of prenatal androgens are responsible for the association between autistic traits and facial appearance, it would be surprising if they had the identical long-lasting effect on female and male facial development, given the sex difference in absolute levels and the contributing factors from other hormones during development. Previous studies of the association between masculinity and testosterone have been largely or entirely limited to males (e.g., Schaefer et al., 2005; Penton-Voak & Chen, 2004; Sisk et al., 2003; Verdonck et al., 1999). Research into the correlations between pre-natal, pubertal and adult testosterone levels in humans would further our understanding of such sex differences.

Our findings are consistent with some previous investigations on the effects of perinatal androgens in humans. These findings are based on the ratio between the lengths of the second and fourth fingers (2D:4D) as a marker of prenatal androgen levels (Manning, Scutt, Wilson, & Levis-Jones, 1998). The 2D:4D ratio has frequently been found to be sexually dimorphic, with the ratio lower in males than females (Manning et al., 1998). Some recent reviews argue that 2D:4D is a specific consequence of pre-natal androgens (Honekopp, Bartholdt, Beier, & Liebert, 2007; McIntyre, 2006), a view which is supported by observations of a stable 2D:4D dimorphism as early as 9 weeks after gestation (Malas, Dogan, Evcil, & Desicoglu, 2006). Although the reliability of 2D:4D as a correlate of masculine traits and behaviours is not always high (Putz, Gaulin, Sporter, & McBurney, 2004), 2D:4D has been previously correlated with ASD. A recent meta-analysis by Teatero and Netley

(2013) found that individuals with ASD had, on average, 2D:4D ratios 0.10 – 0.77 standard deviations lower than controls. Additionally, systemising scores were negatively correlated with 2D:4D, whilst empathising scores showed a positive correlation. In this context of 2D:4D as an early androgen marker, it is therefore relevant that adult 2D:4D is negatively correlated with adult facial masculinity in adult men (Neave, Laing, Fink, & Manning, 2003), and using the same database, facial shape (Schaefer et al., 2005). This correlation is even present in prepubertal boys, consistent with a possible direct influence of prenatal androgens on facial masculinity (Meindl, Windhager, Wallner, & Schaefer, 2012).

Our findings do not agree with the conclusions of Bejerot et al. (2012), who concluded that ASD was not characterized by masculinisation of facial features. There are participant differences between the studies: Bejerot et al. compared ASD and controls rather than high and low-AQ; and used only 8 observers. However, we suggest the important difference is the judgement task for observers: we asked observers to judge sexual dimorphism, and Bejerot et al. asked for judgements of sex typicality-coherence. Bejerot et al. used a scale in which high sex-typicality was at one extreme, low sex-coherence at the other extreme, and average typicality-coherence in the middle. Bejerot et al. did not further elaborate upon or describe the terms of their scale; however, we would note that, as discussed earlier, sexual dimorphism is not the same as sex-typicality, certainly not in the statistical sense. High levels of sexual dimorphism are, by statistical definition, not typical; and highly typical faces for a given sex are, by statistical definition, not the most dimorphic. We therefore suggest a typicality-coherence scale is not well suited for assessing issues relating strictly to dimorphism, as extreme dimorphism would be rated as lower in typicality than average dimorphism. By this interpretation, the faces of male controls in Bejerot et al. were (nonsignificantly) more typical, not more dimorphic, than those of the ASD males. Similar reasoning applies to their results on body typicality-coherence. In contrast, our measure directly addresses issues relating to sexual dimorphism, as we simply asked observers to choose the more masculine of two faces, one based on the statistical qualities of high-AQ scorers, and one based on the statistical qualities of low-AQ scorers.

As highlighted by Hammond (2008) the phenotype and aetiology of ASD are heterogeneous, resulting in problems in identifying the cause and developing effective treatments for the disorder. Hammond argues that analysis of the ASD phenotype could help to identify homogenous subgroups, which could inform a more incisive analysis of the genes and mechanisms involved in its pathogenesis. Our results suggest that phenotypes related to facial masculinity may be an interesting line of exploration here. That is, facial masculinity might be useful in parcelling a more homogenous subgroup, of individuals who hold a specific phenotype, from a heterogeneous population of ASD issues.

Facial masculinity in men is an important social cue, and our findings suggest a relatively unexplored possibility of a mismatch between masculine appearance and behaviour in high AQ men. Studies have had mixed findings in terms of the social implications of perceptions of masculinity, as reviewed by Scott, Clarke, Boothroyd and Penton-Voak (2012). Scott et al. highlight the lack of consistent evidence for popular theories of facial masculinity, such as the immunocompetency theory, where facial masculinity is proposed to signal high genetic fitness due to the ability to overcome the negative immunosuppressive effects of testosterone. As an alternative, Scott et al. propose female preferences for masculinity reflect preferences for competitive mates, that is, more masculine males are more likely to rate higher in intrasexual competition due to aggressiveness and dominance. These masculine traits of aggression and dominance are not necessarily characteristic of high AQ. There may therefore be a mismatch between expectancies of observers based on facial masculinity, and the actual behavior of high AQ men. This mismatch, with individuals with high AQ scores being perceived to have social traits that they do not exhibit could lead to social problems.

The results from this study could be applied to theories further afield from the extreme male brain theory, such as the diametric hypothesis from Crespi, Stead and Elliot (2010). Here it is proposed that ASD and psychotic spectrum disorders such as schizophrenia and bipolar disorder form the opposing ends of a single continuum. Additionally it is suggested that ASD is the consequence of paternal genetic imprinting, whereas psychotic spectrum disorders are caused by maternal imprinting (Crespi & Badcock, 2008). Taking into account our findings that high AQ males are perceived to be more masculine in facial appearance, further research could be conducted to see if individuals with schizophrenia spectrum disorders are perceived to be more feminine in appearance – developing support for the combination of these theories.

The finding that high AQ male faces had features that were consistently perceived as more masculine lends some support for the extreme male brain theory (Baron-Cohen et al., 2002). That is, for men, there seems to be a correlation between appearance and behaviour, which could plausibly be explained by hormonal factors. However it must be noted that a student, not a clinical, sample was used and scores were generally below the proposed cut-off for clinically relevant traits. For this reason our findings should be taken, first for their potential insight into issues surrounding high AQ, including a possible mismatch of expectations and behaviour regarding masculinity; and second, as a justification for further research looking at facial masculinity in clinical samples of

ASD; to better develop understanding into the possible social implications of observer impressions of masculinity in ASD, particularly in adults.

The next step

This chapter lends support for the extreme male brain theory, demonstrating a masculinised phenotype in males with high AQ scores. Badcock and Crespi (2008) detail an associated theory whereby ASD sits on the extreme, masculine, end of a gender spectrum with psychotic disorders such as bipolar disorder and SPD sit at the other, feminine, extreme. The next chapter seeks to determine if the positive results found for ASD in this chapter are extended to the perception of gender in psychotic disorders.

Chapter eight: Facial dimorphism and the diametric hypothesis

Abstract

The diametric hypothesis (Crespi & Badcock, 2008) proposes that autistic spectrum and psychotic disorders form two extremes of a mechanistic - mentalistic spectrum. This theory suggests these disorders are respective representations of paternally and maternally imprinted genes, suggesting ASD is highly masculinised whilst psychotic disorders femininised. This study aimed to determine if these traits were represented in observers perceptions of facial characteristics. Stimuli were created from high and low female symptom scorers of the "psychotic spectrum" disorders; schizotypal personality disorder, borderline personality disorder, eating disorders, depression and bipolar disorder and high and low male scorers of the Autistic Quotient. In a two-alternative-forced-choice study observers were asked to pick which face they felt was more feminine or masculine in appearance. Depression and eating disorders were found to be perceived as more feminine in appearance, whereas ASD and borderline personality disorder were perceived to be more masculine. This study lends support to the concept of a continuum of masculinity and femininity in common mental health disorders.

8.1.Introduction

The extreme male brain theory of autism

Baron-Cohen's (2002) extreme male brain theory of autism proposes that systemising and empathising are two extremes of a continuum of cognitive styles, with normality lying in the middle. The theory proposes that these extremes can be thought of in terms of sex typicality, with systemising being a predominantly male trait and empathising female. Baron-Cohen draws evidence for the superiority of systemising skills in men from a variety of examples, such as the systematic disciplines of maths, physics and engineering being largely male dominated, boys out performing girls in constructional tasks from an early age and the apparent male superiority in attending to relevant details. Conversely, Baron-Cohen argues females are superior in empathising tasks such as sharing, responding to the distress of others and showing sensitivity to non-verbal communication such as facial expressions (Baron-Cohen, 2002).

Baron-Cohen (2002) proposes that autistic spectrum disorders (ASD) are a consequence of an "extreme" male brain, characterised by extreme systemising behaviour. This extreme systematic brain type is proposed to be a consequence of over exposure to testosterone in pre-natal development, a theory for which Baron-Cohen has amassed a wealth of evidence. Firstly, foetal androgen tests have demonstrated a correlation between increased androgen levels in amniotic fluid samples and subsequent ASD diagnoses (Auyeung et al., 2009; Ruta, Ingudomnukul, Taylor, Chakrabarti, & Baron-Cohen, 2011). This is further supported by the high male:female ratio found in ASD, with males three to five times more likely to be diagnosed than females (Bailey et al., 1995; Cosgrove & Riddle, 2004; Fombonne, 1999; Rodier, Bryson, & Welch, 1997). A ratio which has been reported to increase when milder forms of ASD are assessed, for example after screening with the Autistic Quotient (AQ, Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 20001) Brugha and colleagues (2011) reported a male:female ratio of 9:1.

Further evidence for the increased pre-natal exposure to testosterone in ASD comes from findings based on the ratio between the lengths of the second and fourth fingers (2D:4D). This has been shown to be a sexually dimorphic trait, with males demonstrating a lower ratio than females (Manning, Dogan, Evcil, & Desdicoglu, 1998) and is thought to be driven by prenatal androgen exposure (Manning, Scutt, Wilson, & Levis-Jones, 1998). Individuals with ASD have consistently been found to have a lower, more masculine, 2D:4D ratio than controls (Teatero & Neatley, 2013), further supporting the proposal that ASD is, at least, correlated with increased pre-natal testosterone levels.

A final strand of support comes from the increased reporting of testosterone-related symptoms by females with ASD, such as irregular menstrual cycles, an increased risk of polycystic ovary syndrome and the increased prevalence of bi- or asexuality (Ingudomnukul, Baron-Cohen, Wheelwright, & Knickmeyer, 2007). Furthermore when completing the Autistic Quotient, females with a genetic disorder which causes the overproduction of foetal androgens, congenital-adrenal hyperplasia, had a significantly larger, more autistic, overall score than controls (Knickmeyer et al., 2006).

The diametric hypothesis

Badcock and Crespi (2006) proposed a new theory to explain the biological underpinnings of the extreme masculinity found in ASD. Here they proposed that, instead of an over-exposure to testosterone in pre-natal development, as suggested by Baron-Cohen (2002), ASD may be the result of a bias towards paternally expressed genes associated with brain development. In 2008 this theory was extended upon to develop the diametric hypothesis (Badcock & Crespi, 2008; Crespi & Badcock, 2008; Crespi, Stead, & Elliot, 2010). Here, Crespi and Badcock propose that ASD and psychotic spectrum disorders, such as depression, schizophrenia and bipolar disorder, have diametrically opposed phenotypes, resulting from alterations in genomic imprinting favouring either paternal or maternal genes respectively.

A key component of this diametric phenotype is social cognition, with the two disorders representing the two extremes of the cognitive spectrum, with normality in the centre. Whereas ASD is characterised by underdeveloped social cognition, it is hyper-developed in psychotic disorders. The apparent converse social cognition skills are highlighted when key symptoms of each disorder are compared. For example, the inability for individuals with ASD to understand the social norms of groups can be compared with the paranoid delusions of individuals with psychotic disorders, who believe there is group conspiracy everywhere. Also, whereas individuals with ASD have chronic deficits in theory of mind, individuals with psychotic disorders often have religious, mystical and magical delusions where they attribute intention and meaning to everything. Finally, ASD is characterised by deficits in gaze, symptoms opposite to the paranoid delusions of being spied on in individuals with psychotic disorders (Badock & Crespi, 2008)

For Badcock (2009) this continuum of social cognition is best described in terms of mechanistic and mentalistic thinking. Mechanistic, ‘things’, thinking is a system of cognition specific to the physical world, a mechanical, mindless way of seeing the world which has evolved for the interaction with the physical world. Conversely mentalistic, ‘people’, thinking is concerned with understanding fellow human beings, their minds, motives and emotions and has evolved to interact with other people in the psychological environment. Crespi and Badcock (2008) go on to suggest that ASD is characterised by mechanistic thinking, whereas psychotic disorders represent the mentalistic end of the cognitive continuum. This proposal is not dissimilar to Baron-Cohen's (2002) systemising-empathising continuum, however Badcock (2009) argues that Baron-Cohen's terms do not reflect the cognitive processes involved in the same manner that the terms mechanistic and mentalistic do.

Crespi and Badcock (2008) propose that this continuum of social cognition, which manifests in the symptoms associated with ASD and psychotic disorders, is a consequence of alterations of genomic imprinting. Whereas in most cases the alleles from both parents are expressed simultaneously, in a small proportion of cases genes are imprinted, that is gene expression only occurs from one allele. In these cases the gene may be maternally imprinted, that is the gene expressed is from the mother, or they may be paternally imprinted, that is the father's gene is expressed. These genes are involved in a developmental, physiological tug-of-war, where each parent gains from having their genes expressed. A bias towards paternally expressed genes result in enhanced growth and a tendency to be demanding and selfish during interactions with the mother, whereas a bias towards maternally expressed genes result in constrained growth and thus smaller babies who are placid and less demanding, all of which are energetically 'cheaper' for the mother (Badcock & Crespi, 2008; Crespi & Badcock, 2008).

Whilst small biases towards maternally or paternally imprinted genes can benefit the mother or father respectively by decreasing and increasing maternal demand, Crespi and Badcock (2008) propose that large alterations in gene expression are the cause of a range of neuropsychiatric disorders. A bias towards paternally expressed genes result in children on the mechanistic end of the cognitive continuum, causing the development of ASD symptoms. Whereas a bias towards maternally expressed genes result in children on the mentalistic end of the continuum, who are likely to exhibit psychotic symptoms. Crespi and Badcock draw support for this theory from multiple domains including diametric findings in brain size, birth weight, growth, neurological function, neurological development, cognitive and behavioural studies (Badcock, 2009; Crespi & Badcock, 2008). One striking piece of evidence for the theory comes from the implications of genomic imprinting on chromosome 15. A paternal gene bias in this area causes Angelman syndrome, characterised by frequent laughter and walking, hyperactivity and prolonged suckling. Whereas a maternal bias in this area causes Prader-Willi syndrome, characterised in infancy by reduced appetite, poor suckling, a weak cry, inactivity and sleepiness prior to weaning - however after infancy the opposite is the case and the child is prone to extreme and unselective overeating. The most striking difference between these two disorders is the prevalence of neuopsychiatric disorders, there is a very high incidence of ASD in children with Angelman syndrome and rates of psychosis with depression are exceptionally high in Prader-Willi syndrome (Badcock & Crespi, 2008; Crespi & Badcock, 2008).

Crespi and Badcock (2008) have demonstrated how a bias towards maternal or paternal imprinted genes can result in children with traits associated with psychosis or ASD respectively, however the degree to which these symptoms are expressed can be further modified by the sex of the child. So far we have discussed the implications of genomic imprinting favouring maternal or paternal genes, but as illustrated in figure 8.1 there is an additional axis to this theory, the expression of sex chromosomes. Crespi and Badcock put forward that there are two axes of social cognition, one which is determined by sex, the other by the maternal or paternal bias in gene expression. The interaction of these two axis can help to overcome short-comings of other theories, such as the inability of the extreme male brain theory to explain why although ASD is less common in females, but when it does occur it often causes severe deficits. Crespi and Badcock suggest that disorders are more common, but less severe when the two axes are compatible, such as when a bias in paternally expressed genes causes Asperger's syndrome in males, or when maternally expressed genes cause Schizotypy in females. However, a paternal bias in females can cause Rett syndrome - a form of ASD with a more equal sex ratio but increased intellectual disability. A maternal bias in males cause Paranoid schizophrenia - a form of schizophrenia with more severe symptoms which is more common in males.

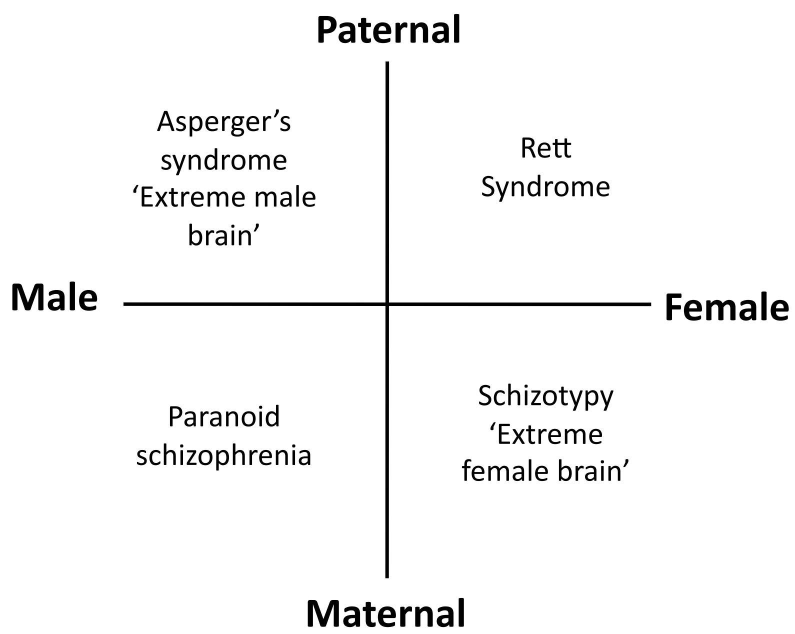


Figure 8.1. Figure detailing the continuums from maternal to paternal imprinting and the effects on males and females (Badcock, 2009).

The extreme female brain

It has been shown how the extreme male brain theory of autism proposes that ASD is a characterised by a extreme systemising, masculine, thought pattern caused by over exposure to testosterone in pre-natal development (Baron-Cohen, 2002). Crespi and Badcock (2008) share some important similarities to the extreme male brain theory, as both theories agree that ASD is characterised by extreme masculinity in cognition.

For this reason, many authors have combined the theories and grouped autistic spectrum disorders as representative of the ‘extreme male brain’ and psychotic disorders as the opposite ‘extreme female brain’. Although it is noted that the two theories propose differing causal basis for the phenotypes associated with ASD and psychotic illnesses, from here on the diametric hypothesis shall be refered to in terms of the extreme male brain, with mechanistic/systemising cognition resulting in phenotypes associated with ASD and the extreme female brain, with phenotypes associated with mentalistic/empathetic cognition resulting in phenotypes associated with psychotic disorders such as depression, bipolar disorder and schizophrenia.

In a test of the extreme female brain’s relationship to empathising and systemising Brosnan, Ashwin, Walker and Donaghue (2010) collected measures of empathising, systemising, anxiety, depression and psychosis. They found, as expected, that empathising and systemising represented diametric concepts of a continuum, and in support of the extreme female brain, psychosis was associated with extreme empathy. However, depression and anxiety were not found to be related to hyper-empathy, as would be hypothesised under the diametric hypothesis. Contradicting these findings other studies have indeed demonstrated a positive correlation between depression and empathy (O’Connor, Berry, Lewis, Mulherin, & Yi, 2007; Zahn-Waxler, Shirtcliff, & Marceau, 2008). Furthermore hyper empathy has also been implicated in eating disorders (Bremser & Gallup, 2012) and borderline personality disorder (BPD, Dinsdale & Crespi, 2013).

The findings that individuals with eating disorders and borderline personality disorder also exhibit the highly empathetic cognitive style demonstrated in the psychotic disorders (schizophrenia, bipolar disorder and depression) grouped by Crespi and Badcock (2008), suggests that this may not be an exclusive group. Here the evidence that a range of disorders, including depression, bipolar disorder, schizophrenia, eating disorder and BPD could be classified as examples of the ‘extreme female brain’ is discussed.

There is evidence for a female bias in disorders of the extreme female brain which parallels the male bias exhibited in ASD. Females are three times more likely than males to be diagnosed with BPD (Widiger & Trull, 1993) and between three and eight time more likely to be diagnosed with eating disorders (Pretti et al., 2009). In an epidemiological study spanning ten countries females were consistently found to have a higher rate of unipolar depression, however the ratio was more equal in bipolar disorder (Weissman et al., 1996). The picture is less clear for schizophrenia, women score higher for positive symptoms of schizophrenia, but men score higher for negative symptoms (Maric, Krabbendam, Vollebergh, de Graat, & Van Os, 2003). These results are paralleled when looking at the less severe end of the schizophrenia spectrum, schizotypal personality disorder (SPD; Raine, 1992).

Another thread of evidence comes from 2D:4D ratios, where-as individuals with ASD are more likely to exhibit a lower, more masculine, 2D:4D ratio, individuals with extreme female brain disorders are more likely to have a higher, more feminine 2D:4D ratio. Evardone, Alexander and Morey (2008) found that a more feminine 2D:4D was a significant predictor of BPD symptoms in both men and women. Similar results were found by Klump and colleagues (2006) demonstrating an association between a more feminine 2D:4D ratio and eating disorder scores in a female sample. In a comparison of male and female schizophrenics and control participants Arato, Frecska, Beck, An and Kiss (2004) found schizophrenics of both sexes to consistently have a more feminine 2D:4D ratio, results replicated by Voracek (2008). Results are less straightforward in depression, with two studies only finding an association between 2D:4D and depression scores in males. Bailey and Hurd (2005) found this association when using the NEO-PI depression subscale in adults and Vermeersch, T'Sjoen, Kaufman and Vincke (2008) found a significant correlation between depression scores and 2D:4D in adolescent boys, but not girls. To date, the authors have been unable to find any reports assessing 2D:4D ratio and bipolar spectrum disorders. A summary of these studies can be found in table 8.2.

Table 8.2. Table detailing findings for masculine and feminine 2D:4D ratios in a range of psychiatric disorders.

Trait	Findings for low (masculine) 2D:4D	Findings for high (feminine) 2D:4D
ASD	Teatero & Neatley (2013)	
BPD		Evardone et al. (2008)
SPD		Arato et al. (2004)
		Voracek (2008)
Eating disorders		Klump et al. (2006)
Bipolar disorder	-	-
Depression		Vermeersch et al. (2008)
		Vincke (2008)

This review has given a brief overview of evidence suggesting that eating disorders and BPD exhibit a similar extreme female brain phenotype associated with the psychotic disorders (bipolar disorder, schizophrenia and depression) grouped by Crespi and Badcock (2008). Not only do these disorders have similar biases in female diagnosis, they are also associated with a more empathetic cognitive style and are consistently found to have a more feminine 2D:4D ratio (however it must be noted that associations between bipolar disorder and 2D:4D have not been reported, reducing the evidence for this particular disorder). This does not just suggest a similarity between the disorders, but it also shows a marked contrast from the extreme male brain spectrum of autistic disorders which are characterised by a male bias in diagnosis, systematic cognitive style and masculine 2D:4D ratio. Although not originally included by Crespi and Badcock (2008) in the range of extreme female brain disorders, additional research conducted has proposed that eating disorders (Bremser & Gallup, 2012) and BPD (Dinsdale, 2008) could also be included within the group.

Sexual dimorphism in the face and cognitive style

Previous research has demonstrated that males with high Autistic Quotient (AQ, Baron-Cohen et al., 2001) scores are more likely to be rated as more masculine in facial appearance than those who scored lowly (Scott, Jones, Kramer, & Ward, 2015). This study aims to further this research by determining if observers perceive facial images of females scoring highly on measures of disorders characterised by the extreme female brain to be more feminine in appearance than those scoring lowly. Having a diametric perception of gender, that is having facial images of the extreme male brain disorder, ASD, rated as more masculine and facial images of disorders of the extreme female brain disorders as more feminine, would lend support for a phenotype associated with the continuum of masculine to feminine proposed by Crespi and Badcock's (2008) diametric hypothesis.

In order to test this hypothesis we collected facial photographs of over 200 university students alongside measures of ASD, eating disorders, schizotypal personality disorder (SPD), BPD, bipolar spectrum disorder and depression. These were then presented to observers who rated them for gender on a likert scale ranging from very feminine in facial appearance to very masculine in facial appearance. In this study only female stimuli for disorders of the extreme female brain, and male faces for the extreme male brain disorder were used, this is because of the complicated interactions between hormone levels in alternate sexes, a topic discussed later.

8.2. Method

This study comprised of two phases, in phase one facial photographs were collected alongside measures of ASD, eating disorders, SPD, BPD, bipolar spectrum disorder and depression. In phase two these were then presented to naive observers who were asked to make judgments about the gender of the stimuli.

8.2.1. Phase one: Stimuli creation

This phase aimed to develop stimuli which reflected differences, if any, between individuals scoring high and low on traits of the extreme male and extreme female brain. To create stimuli for this study facial photographs alongside measures of each trait were collected. The photographs of the highest and lowest scorers for each trait were 'averaged' together, to create composite images. The photographs were collected at two time points, in photo shoot B (See Appendix B for details) measures of SPD, bipolar spectrum disorder, depression and ASD were collected. During photo shoot C measures of eating disorders and BPD were collected, with a different sample.

8.2.1.1. Participants

Collection one: Two hundred and twenty one (130 females) Bangor University students (Age M = 21.65, SD = 5.09) took part in photo shoot B (See appendix B for details) and and paid £6 for their participation.

Collection two: Two hundred and forty (141 females) Bangor university students (Age M = 21.69, SD = 5.19) took part in photo shoot C (See appendix B for details) and paid £6 for their participation.

8.2.1.2. Measures

Schizotypal personality disorder: SPD was measured using the Schizotypal Personality questionnaires (Raine, 1991). This is a 72 item questionnaire where symptoms are identified using a "Yes/No" criteria with a possible score range of 0-72. Female scores ranged from 0-67, *M* = 24.10, *SD* = 15.09.

Bipolar spectrum disorder: Bipolar disorder was measured using the Mood disorder questionnaire (Hirschfeld et al., 2000). This is a three-part questionnaire, in part one participants are presented with 13 statements and asked to indicate if they have ever experienced the symptoms. Each answer of 'yes' is scored one point, which is then summed. Female participant scores ranged from 0-13, *M* = 7.64, *SD* = 3.22. In the second part participants are asked if they have experienced several of these symptoms at the same time, 80 female participants indicated this had happened to them. Finally participants are asked how much of a problem these symptoms had caused to them, ranging on a four-point scale from no problem to serious problems. Twenty female participants indicated that their symptoms had caused them moderate to serious problems. In order to classify as having significant bipolar spectrum symptoms participants must score seven or higher in part one, indicate that they had experienced multiple symptoms at one time and that they had caused moderate to serious problems in their lives. In this sample 11 female participants met this criteria.

Depression: Depression was measured using the inventory of depressive symptoms (Rush, Gullion, Basco, Jarrett, & Trivedi, 1996). This is a 30-item questionnaire where participants are asked to describe symptom severity on a scale of 0, no symptoms at all, to 3, high symptom presence. Scores from each question are summed resulting in a total possible score of 84. Female scores ranged from 1-62, *M* = 15.41, *SD* = 11.56.

Autistic spectrum disorder: ASD was measured using the Autistic quotient (AQ, Baron-Cohen, Wheelwirhgt, Skinner, Martin, & Clubley, 2001). This is a 50-item questionnaire which assesses subscales of ASD traits; Attention to detail, attention switching, social skills, communication and imagination. Each question is answered on a four point scale from definitely agree to definitely disagree; a point is given for the extreme answer (definitely agree/disagree) relating to ASD traits. Male scores ranged from 3 - 30, *M* = 17.74, *SD* = 5.76.

Eating disorders: Symptoms of eating disorders were measured using the 26-item Eating attitudes test (EAT-26, Garner, Olmsted, Bohr, & Garfinkel, 1982). In this questionnaire participants are asked to read 26 statements about eating behaviours and indicate how often they exhibit them on a scale of Always - Never. The maximum score for each question is 3, resulting in a top possible score of 78. Female scores ranged from 1- 62, *M* = 11.58, *SD* = 12.38.

Borderline personality disorder: Symptoms of BPD were measured using the 23-item borderline symptom list (BSL-23, Wolf et al., 2007). In this questionnaire participants are presented with 23 problems for which they indicate how much they suffered from over the previous week on a scale of not at all (0) to very strongly (4). Overall score is computed by summing all items resulting in a top possible score of 92. Female scores ranged from 0 -78, *M* = 16.53, *SD* = 15.49.

8.2.1.3. Procedure

Participants were asked to remove all make-up and jewellery, tie back their hair. Photographs were taken against a light background whilst participants facing the camera holding a neutral facial position. Photographs were taken from a distance of 2m, camera height, zoom and flash were kept constant throughout. Individuals who did not report themselves to be white-British, were aged over 30 and men with facial hair were removed from the sample.

To create the stimuli the highest and lowest 18 females for SPD (high *M* = 46.17, *SD* = 8.02, low *M* = 4.83, *SD* = 2.75), bipolar disorder (high *M* =10.78, *SD* =

2.34, low $M = 3.94$, $SD = 1.89$), depression (high $M = 32.39$, $SD = 7.06$, low M 4.67, $SD = 1.71$), eating disorders (high $M = 35.61$, $SD = 13.52$, low $M = 0.94$, $SD = 0.64$) and BPD (high $M = 45.72$, $SD = 19.05$, low $M = 1.89$, $SD = 1.28$) and the and the highest and lowest 18 males for ASD (high $M = 24.17$, $SD = 2.75$, low $M = 10.50$, $SD = 2.62$) were selected. These image groups were then randomly allocated to a balanced design where nine composite images were created by 'morphing' together four individual images (See Appendix A). Due to the balanced design individual images only appeared in two composites, and there were never more than two of the same individuals in any one composite for each disorder. This method was adopted to ensure that any effects found were not being driven by a small number of individuals in a group. This process resulted in the creation of 108 composite images, nine high and nine low for each trait, examples of which are depicted in figure 8.3.

Trait	High	Low
ASD		
BPD		
Bipolar		
Depression		
Eating disorders		
SPD		

Figure 8.3. Figure showing examples of stimuli created from the high and low scorers on each mental health trait.

8.3. Phase two: Observer perceptions of gender

In phase one 108 composite images representing the facial appearance of individuals scoring high and low on six traits associated with the extreme male and female brain. In this phase these images were presented to naïve observers who rated them for gender.

8.3.1. Observers: Thirty-four (28 female) Bangor university students (age *M* = 21.41, *SD* = 5.51) were compensated with course credits for their participation.

8.3.2. Procedure: Each of the 108 composite images were presented individually, in a random order, to observers who were asked to rate the image on a Likert scale of 1 (very feminine) to 6 (very masculine) by using the mouse to click on the appropriate number. Responses were not speeded and viewing distance was not fixed. In this experiment a Likert rating methodology was utilised as it allowed for comparison across traits, analysis not applicable to 2AFC methodology.

8.4. Results

Mean scores were calculated for the nine high and nine low composite images for each trait. A 6 (trait) x 2 (symptom severity) x 2 (observer sex) repeated measures ANOVA showed a significant effect of trait, $F(2.85, 91.16) = 3.71$, $p = .016$, of symptom severity, $F(1, 32) = 16.72$, $p < .001$ and an non-significant effect of observer sex, $F(1, 31) = 0.005$, $p = .945$. Paired samples t-tests were then used to determine if there was a significant difference in gender rating between high and low scorers. A lower score indicates a more feminine facial appearance, whereas a higher score indicated a more masculine appearance.

High depression images were rated as significantly more feminine in facial appearance than low scorers (2.52 vs 3.29, $t(33) = 9.35$, $p < .000001$, $d = 1.08$, 95% CI [-.94, .60]). Similar results were found with eating disorders, again a more feminine facial appearance was attributed to higher scorers (2.68 vs 2.92, $t(33) = 2.93$, $p = .006$, *d* = .50, 95% CI [-.40, -.07]). Although trending in the same direction results for SPQ were only found to be marginally significant (2.63 vs 2.72, $t(33) = 1.76$, $p = .088$, $d = .29$, 95% CI [-.20, .01]). Conversely females with high BPD scores were rated as significantly less feminine in facial appearance than those who scored lowly (2.86 vs 2.65, $t(33) = 3.37$, $p = .0019$, $d = .58$, 95% CI [.09, .35]). Results for bipolar disorder also failed to reach significance (2.71 vs 2.61, $t(33) = 1.55$, $p = .13$, $d = .27$, 95% CI [-.03, .24]). Finally, males with high AQ scores were rated as significantly more masculine in facial appearance than those scoring low (4.69 vs 4.36, $t(33) = 4.16$, $p = .0002$, $d = .72$, 95% CI [.17, .49]), replicating previous work (Scott et al., 2015). [11]

Figure 8.4 illustrates the differences between scores attributed to high and low images of each trait. As depicted images of individuals scoring highly for depression and eating disorder were perceived to be significantly more feminine in appearance than low. Difference scores for SPD and bipolar spectrum disorder grouped around the zero mark, suggesting no difference in appearance, and images with high degrees of BPD and ASD symptoms were perceived to be more masculine in facial appearance.

---

[11] Please note, in this study correcting for multiple comparisons did not reduce the significance of the findings

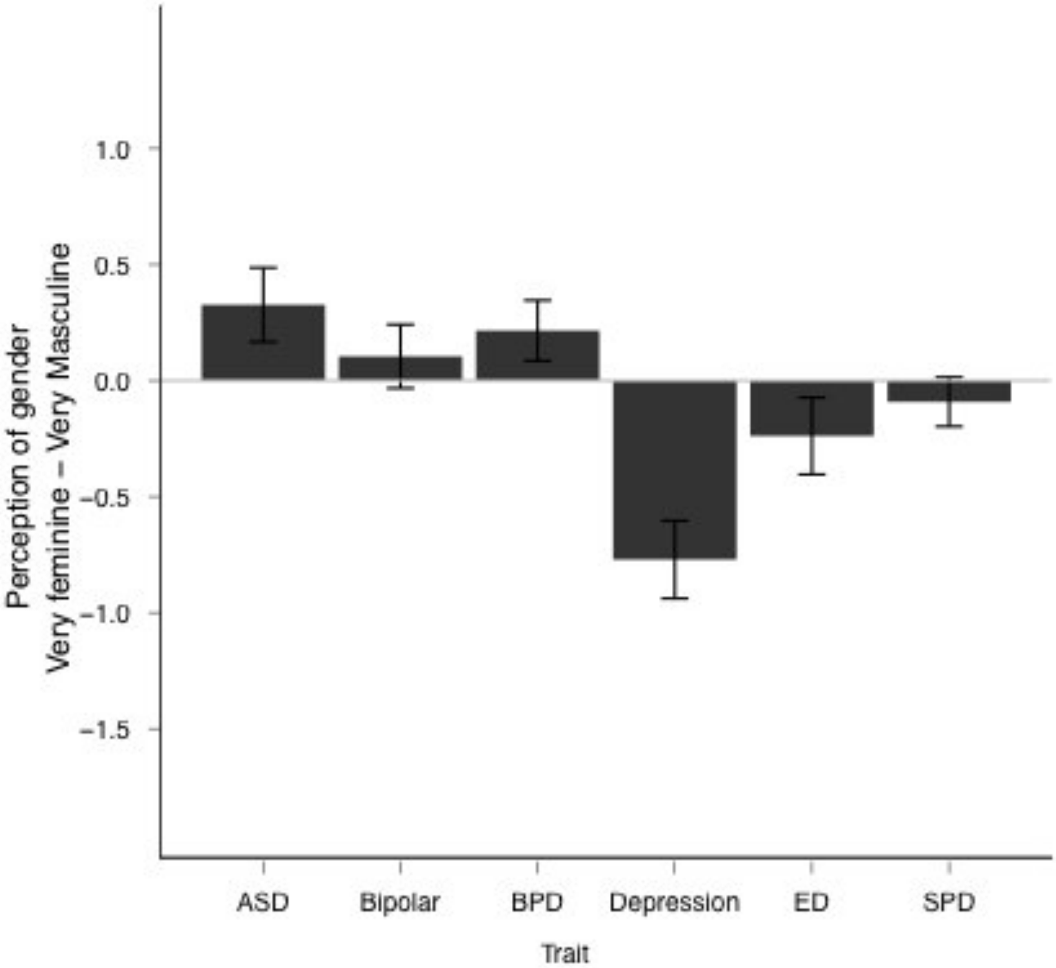


Figure 8.4. Figure detailing perceptions of gender across each of the mental health traits assessed, with results below 0 indicating a more feminine appearance and those above 0 indicating a more masculine appearance. Error bars represent 95% confidence intervals from the mean.

8.5. Discussion

These findings lend some support to a diametrically opposed phenotype associated with disorders of the extreme male and female brains. It was demonstrated that images of females scoring high on measures of depression and eating disorder symptoms were perceived to be more feminine in facial appearance than those scoring lowly. Results for bipolar spectrum disorder and schizotypal personality disorder failed to reach significance and findings regarding borderline personality disorder were opposite to hypothesised, with faces of individuals with high degrees of traits being rated as more masculine in facial appearance. Facial images of males scoring high on the autistic quotient were rated as significantly more masculine in appearance than those scoring low, replicating findings from Scott and colleagues (2015) using a new stimuli creation procedure.

In general these findings lend a degree of support to Crespi and Badcock's (2008) notion that ASD and psychotic disorders being diametrically opposed. That is, a marked difference in the perception of gender in individuals with high symptoms of ASD than that found in a number of the psychotic spectrum disorders was found. However, results for two of the key disorders proposed by Crespi and Badcock to form the psychotic end of the diametric hypothesis, schizophrenia and bipolar disorder failed to reach significance. That is, there was no significant perceived differences in gender between individuals with high and low traits. Here the possible reasons for these findings are discussed.

Crespi and Badcock (2008) include schizophrenia within their umbrella of psychotic disorders, and whilst this study collected a measure of schizotypal personality disorder this should not infringe upon these results negatively as it is considered to be a disorder on the schizophrenia spectrum (Battaglia & Torgesen, 1996) and Crespi and Badcock highlight the potential use of sub-clinical samples in the testing of their theory (Crespi & Badcock, 2008). One cause of the non-significant results found could be the use of the full schizotypal personality questionnaire (Raine, 1991). Due to the skewed gender ratio found in schizophrenia, and indeed schizotypal personality disorder, where negative symptoms are more common in males and positive symptoms more common in females (Maric et al., 2003; Raine, 1992) Crespi and Badock limit their theory to the inclusion of positive symptoms of Schizophrenia. The use of a full SPD scale could have an effect on these results, and future study could benefit from replication just using the sub-scale assessing the positive symptoms of SPD to determine if this is the case.

Another disorder which failed to exhibit a significant difference in this study was bipolar spectrum disorder. The suggested reasoning for this non-significant finding is the strict criteria associated with the mood disorder questionnaire. In order to fulfill the criteria of this questionnaire individuals must indicate that they have experienced seven or more of the clinical symptoms, that they have experienced them at the same time and that they have caused a significant impairment on their life. Out of the 130 females who took part in the photo shoot only 11 fulfilled this criteria, short of the 18 individuals required to develop the composite stimuli. This meant that the remaining seven individuals' symptoms do not match that required for further investigation. Replication of the study with a clinical sample could determine if this indeed the reason for our non-significant findings. Another possibility lies within the lack of evidence for a female bias in bipolar disorder, indeed female:male ratio of bipolar disorder is equal (Weissman et al., 1996) and there is a dearth of evidence suggesting a female bias in 2D:4D ratio. This suggests that the story for bipolar disorder is not as straightforward, however future research assessing the difference in female bias when comparing patients with bipolar disorder who do and do not have psychotic symptoms may give more insight. This is highlighted by the association between empathy and psychotic symptoms demonstrated by Brosnan and collegues (2010) rather that bipolar disorder per-sae.

The final set of results which did not follow the hypothesis was results of BPD. This disorder was not included in Crespi and Badcock (2008)'s psychotic spectrum, but rather proposed in later work supported by Crespi (Dinsdale, 2008). The foundations for inclusion of BPD appear sound, with a high female:male ratio and evidence suggesting increased pre-natal oestrogen levels via 2D:4D ratio's. One reasoning behind the reversal of gender attributions may lie in the perceptions of females with BPD. In previous chapters I have demonstrated the ability for naïve observers to be able to accurately differentiate between the stimuli used in this study. That is, observers were able to accurately determine between the stimuli in terms of which image was created from individuals scoring high and low on the BSL-23. However, when asked to rate the images for attractiveness, results for females failed to reach significance. As perceptions of femininity and attractiveness are highly correlated, the lack of a significant finding with perceptions of attractiveness could cause a bias in perceptions of femininity.

The research discussed here is predominantly based on evidence from pre-natal levels of androgen and oestrogen exposure, which does not have a proven direct impact on adult face shape. A popular hypothesis explaining the association between pre-natal and pubertal sex hormone levels is the long-standing organisational-activation model (Phoenix et al., 1959). This theory proposes that pre-natal sex hormone levels serve to 'organise' brain structures, which are then 'activated' by sex hormones during puberty. Although a large majority of the work into prenatal and pubertal sex hormones has been conducted into male androgens, it would be plausible to suggest that a similar mechanism could illustrate a correlation between pre-natal and pubertal oestrogen levels in females. At puberty oestrogen protects females against the masculinising effects of androgens, resulting in the classic female face shape. So if the organisational-activational hypothesis was to be adopted as a method of explaining the biological underpinnings of the diametric phenotypes of the extreme male and female disorders, an over exposure to testosterone and oestrogen respectively during pre-natal development could 'organise' the brain in a highly masculine or feminine way, which would be activated at puberty, when facial dimorphism occurs. In this experiment we only used male faces in the study of ASD, and only female faces in the study of BPD, SPD, depression, bipolar spectrum disorder and eating disorders. This is as the organisational effect of pre-natal sex hormones on opposite sexes, e.g.the lasting effect of oestrogen in men, has had little research and the effects are likely to be less straightforward across sexes.

This hormonal basis for the dimorphism in facial shape correlates with the biological underpinnings proposed by Baron-Cohen (2002), however Crespi and Badcock (2008) adopt the explanation of genomic imprinting rather than hormone influences. Thus the potential implications of the organisational-activational model are more of an extension of Baron-Cohen's extreme male brain theory, utilising the phenotypic associations made by the diametric hypothesis.

These findings demonstrate a significant difference in perceptions of gender when individuals score high or low on measures of ASD, depression and eating disorders. This could have implications for social situations, as individuals perceive females with a very feminine appearance to be warmer, more honest and severe (Berry & McArthur, 1985) and males who are very masculine in appearance to be aggressive and dominant (reviewed by, Scott, Clark, Boothroyd, & Penton-Voak, 2012). The effect of these perceptions could also be diametric. For women, perceptions of femininity attribute positive social connotations which are likely to increase social interactions, however the perceptions of aggression and dominance in males is likely to do the reverse. Additionally there is the potential effects of perceived dishonesty, as males with high AQ scores are unlikely to act in a dominant manner, developing a mismatch between perceived and actual personality traits which may lead to problems in social interactions.

These results indicate a facial dimorphism in gender perception between disorders of the 'extreme female' and 'extreme male' brain. Females reporting high symptom occurrence on measures of depression and eating disorders were rated as significantly more feminine in appearance than those who reported low symptom occurrence. Additionally males with high AQ scores were rated as more masculine in appearance than those scoring lowly. These perceptions could have social implications for individuals, as well as lending support for the diametric hypothesis proposed by Crespi and Badcock (2008). However, the biological underpinnings may fit better with Baron-Cohen's (2002) proposal of later implications of pre-natal sex hormone exposure. It must be noted that these findings utilised a student sample, and although the findings give promising insights to both the potential social implications of these disorders and the impact on existing neuropsychiatric theories, further investigation with clinical samples would be beneficial to clarify the findings.

Chapter nine: General discussion

In this section I recap the general findings from the previous experimental chapters and discuss how they relate to previous studies as well as their potential clinical implications.

Accuracy in perceptions of mental health: A kernel of truth?

Through chapters two to six I investigated the ability of naïve observers to detect a series of common mental health disorders from static facial images, with many positive findings, a summary of which can be found in table 9.1.

In chapter two I demonstrate the accurate discrimination between facial images warped towards the morphological and textural properties common in individuals with high and low levels of depressive symptoms, and in chapter three I demonstrate this ability is still prevalent when the stimuli is reduced down to only show the internal facial features and even when just the eye and brow region is presented. These results suggest that observers need only a very limited degree of visual information to accurately categorise depressive symptoms, much less than utilised in thin-slice video studies, such as Waxer (1976), discussed earlier. These findings most strongly link with Kleinman and Rule's (2013) discovery that observers were able to accurately categorise individuals who were to later commit suicide, and those who would not, from year-book photos. Interestingly within their study they also demonstrated significant results when faces were cropped to only show the internal facial features. However, a recent paper has highlighted some potential methodological issues with studies that utilise stimuli which has been selected to represent an individual in a specific situation. Todorov and Porter (2014) identified a bias when facial images were selected for a specific purpose, e.g. photos most commonly chosen as most appropriate for applying

Trait assessed	Chapter/ study number	Stimuli used	Accuracy	*p*	Effect size	C.I.
Depression	2.1	Full faces	61%	< .001	*d = 1.19*	60-63%
Depression	2.2	Full faces	58%	< .001	*d* =.56	54-62%
Depression	3.1	Internal features	58%	<.0001	*d* = 1.12	55-61%
Depression	3.2	Eye region	56%	.001	*d* = .65	52-59%
SPD	4.1	Male full faces	Male raters - 76% Female raters - 82%	.0027 < .001	*d* = 1.10 *d* = 3.08	61-90% 77-86%
SPD	4.1	Female full faces	Male raters - 48% Female raters - 58%	.68 .027	*d* = .12 *d* = .49	38-58% 51-66%
SPD	4.2	Full male faces	59%	<.0001	*d* = 1.11	56-62%
SPD	4.2	Full female faces	72%	<.0001	*d* = 1.50	66-77%
BPD	5.1	Full male faces	53%	.27	*d* = .21	48-58%
BPD	5.1	Full female face	57%	.00018	*d* = .77	54-61%
ASD	6.1	Full male faces	44%	.028	*d* = .35	39-49%
ASD	6.1	Full female faces	55%	.028	*d = .37*	51-60%
ASD Social skills	6.1	Full male faces	50%	.95	*d* = .01	46-54%
ASD Social skills	6.1	Full female faces	46.5%	.04	*d* = .34	43-50%
ASD Communication problems	6.1	Full male faces	57%	.014	*d* = .42	51-62%
ASD communication problems	6.1	Full female faces	67%	< .001	*d* = 1.13	62-72%

Table 9.1. Table giving an overview of results throughout chapters 2-8

for a high-salary position, were later chosen more frequently as an individual suitable for hire by a second set of participants. These findings suggest that individuals choosing photographs for online sites, such as the faces of suicide website utilised in Kleinman and Rule's (2013) study, are not randomly selected and thus may hold additional cues for observers. Despite the possible stimuli flaws in Kleinman and Rule's (2013) study, paired with my findings, it suggests that there are cues for depressive and suicidal symptoms found within the internal facial features, especially the eye and brow region.

In chapters four and five I extend these findings to explore the possibility that these cues are also observable in schizotypal and borderline personality disorders. Mixed results were established for schizotypal personality disorder (SPD) with a difference in accuracy from female and male observers found in study one. However, as study 2 used a more vigorous stimuli creation methodology, utilising a balanced design for stimuli creation, and a more simplistic and well-authenticated symptom measure, the significant findings of accuracy in both sexes is upheld. In chapter five, assessing accuracy in borderline personality disorder (BPD), significant results were only obtained for female images. To my knowledge there has been no published research assessing the perception of BPD or SPD from facial images, however studies of similar personality disorders have yielded significant results in line with these. Holtzman (2011) developed composites using a similar methodology to those in my research, and found the accurate classification of high/low symptomatology in a triad of personality disorders; psychoticism, narcissism and Machiavellianism. Interestingly, Holtzman also found an increased accuracy for the classification of female faces and when results were separated by sex, findings for males only just tipped into statistical significance. However, as with my stimuli, the sample pool of male faces was much more limited than that for female faces, this reduction in variance could reduce the effect size of results with male images, with results possibly representing a methodological weakness rather than a reduced ability to discriminate traits in males. Viewed as a whole, these findings suggest that the static face does hold cues to personality disorders, but further research with an equal sized stimuli set would be required to determine if these are more readily observed and utilised in female faces.

Finally in chapter six I turn to assess a common neurodevelopmental disorder, autistic spectrum disorder (ASD). Here I demonstrated the accurate discrimination between high and low scores for ASD in females, but the inaccurate projection of symptoms to males with low scores. These results are not entirely consistent with associated previous work by Grossman (2014). Here, observers rated children with high functioning autism as more socially awkward from still video images. Although these results do not lend to the assumption of the recognition of ASD from still images, they do suggest that there are cues to social skills, a prominent deficit in the disorder. Interestingly, in my study I found the highest accuracy in the sub-category of social skills. However, results from the trait assumptions in chapter six may give a likely reason for the inaccurate attributions of symptoms in males - a finding I will discuss later.

These five chapters lend support for the ‘Kernel of truth’ hypothesis (Berry & Wero, 1993) in mental health disorders. That is, they provide evidence for the existence of facial cues to common mental health disorders, which can be utilised by naïve observers to accurately categorise individuals who do and do not have symptoms of these illnesses. But what cues are being utilised to make these decisions?

What holds the cues?

In chapter three I demonstrated that observers were still able to accurately categorise high and low depressive symptom scorers when stimuli were reduced to just show the eye and brow region. Although in the current thesis this procedure was only utilised with the depressive stimuli, other studies have demonstrated accurate trait recognition using similar methods. Over a series of studies Rule and colleagues have demonstrated a maintained level of accuracy when stimuli is reduced in this manner, including studies on male sexual orientation (Rule, Ambady, Adams, & Macrae, 2008, although note again the use of non-random stimuli through dating websites) and female sexual orientation (Rule, Ambady, & Hallett, 2009). However, significant results were not found in relation to religious beliefs (Rule & Ambady, 2010). Furthermore Santos and Young (2011) demonstrate more than 60% agreement between observers on ratings of intelligence and trustworthiness when just presented with the eye and brow region of an individuals face.

Although this sheds some light on where in the face these cues may be held, it does not specify what the cue is. One possibility is symmetry, a cue which has been associated with ratings of facial attractiveness (Gangestad, Thornhill, & Yeo, 1994) and is thought to be an indicator of an organisms developmental stability (Moller & Swaddle, 1997). Indeed Shackelford and Larsen (1997) found individuals who were more facially asymmetrical not only self- reported more physiological, psychological and affective problems but were also rated by observers to be higher in each domain and Hennessy et al. (2004) found more asymmetries in individuals with schizophrenia than controls. These findings could also be linked to the previously discussed findings around minor physical anomalies (MPAs), that is, individuals with a range of mental health disorders have been found to have a higher prevalence of MPAs which could form a basis for a reduction in facial symmetry (although I note studies have failed to find a significant result for individuals with unipolar depression). The social implications of these findings are exemplified when the personality attributions associated with these results are taken into account, as discussed next.

Personality attributions in mental health - the negative stereotype?

As well as the assessment of mental health trait accuracy in chapters two to six, observers were asked to make attributions of personality traits. The findings of each of these studies are collated in table 9.2. There does appear to be a common theme in the trait attribution of high scorers in each of the studies, namely low levels of Agreeableness, Extraversion and Openness alongside higher scores for Neuroticism. These are compounded into a general sense of low social desirability in the ratings of SPD in chapter four. Findings for Conscientiousness were generally non-significant , although when significant (in the attributions of depression and male SPD) they were indeed also negative. It is not unsurprising that perceptions of Conscientious rarely reached significance, as a majority of studies into facial recognition of personality traits have failed to demonstrate an ability to accurately discriminate the trait (Jones, Kramer, & Ward, 2012; Kramer & Ward, 2010; Penton-Voak, Pound, Little, & Perrett, 2006).

Table 9.2. Table detailing the personality attributions given to each psychiatric disorder in chapters 2-8.

Trait	Sex	A	C	E	O	N	Attract	Fem	Social des	Trust
Depression		↓	↓	↓	↓	↑	↑	↑		
Internal features	Male			↓	↓					
	Female									
Eyes	Male			↓	↓					
	Female									
SPD 1	Male	↓	↓	↓	↓	↑	↓			↓
	Female	↑								↑
SPD 2	Male						↓	↑	↓	
	Female						↓	↓	↓	
BPD	Male	↓		↓			↓			
	Female	↓		↓	↓					
ASD	Male			↑			↑	↓		
	Female	↓		↓		↑				

↓ = Undesirable personality traits attributed to high symptom scorers

↑ = Desirable personality traits attributed to high symptom scorers

It was originally proposed that Neuroticism could be a driving force behind the recognition of mental health disorders, due to the strong correlation between the two entities. However, as demonstrated in Table 9.2. high levels of Neuroticism were only attributed to high symptom scorers of depression; males with SPD and females with ASD. One connection that is continuously demonstrated is a negative association between Extraversion and high symptom scorers. That is, low levels of Extraversion were attributed to individuals of both sexes with high scores in depression, males with SPD and BPD and females with ASD. Interestingly, these results are also highly correlated with trait accuracy, so when observers attribute low levels of Extraversion to high scorers of mental health symptoms they are also accurate in the categorisation of those symptoms. One study that highlights this is the misattribution of ASD symptoms to low AQ scoring males in chapter six. Here, observer's ratings of Extraversion and ASD were highly negatively correlated; in that they commonly attributed fewer symptoms of ASD to individuals they perceived to be high in Extraversion. However, possibly due to the perception of masculinity in high ASD scorers, they misattributed high levels of Extraversion to high ASD scorers - causing an inaccuracy in their categorisation of ASD symptoms. There may be a 'kernel of truth' for these attributions; in a recent meta-analysis of studies assessing correlations between the Big 5 personality traits and mental health traits Kotov, Gamez, Schmidt and Watson (2010) found the commonly reported association between mental health disorders and Neuroticism, but also found that many disorders were also associated with low levels of Extraversion. Results replicated by Bunevicius, Katkute and Bunevicius (2008). Indeed, as shown in Appendix B, generally within each photo shoot mental health traits were negatively correlated with Extraversion and positively with Neuroticism, with the exception of bipolar disorder, which exhibited a positive correlation with both traits.

Interestingly, high scores for Extraversion have been associated with facial symmetry, linking back to a possible basis for the recognition of mental health symptoms. The previously discussed study by Shackelford and Larsen (1997) also assessed the association between facial symmetry and Extraversion, finding a positive correlation in females; however results were not replicated in males. In a more recent study, Fink, Neave, Manning and Grammer (2005) found Extraversion to be positively, and Neuroticism to be negatively, correlated with horizontal facial symmetry. Following this, and utilising a more robust methodology and larger sample size, Pound, Penton-Voak and Brown (2007) also demonstrated a significant positive correlation between Extraversion and facial symmetry.

Another trait consistently positively associated with facial symmetry is perceptions of attractiveness (Gangestad, Thornhill, & Yeo, 2004; Thornhill & Gangested, 1999). Again, this is a trait that I found, in general, to be negatively associated with high symptom scorers. As shown in table 9.2, low symptom scorers were perceived to be more attractive in SPD, and BPD. Interestingly high scorers for depressive symptoms were seen as more attractive, although correlation analysis showed that ratings of attractiveness and depression were not significantly related. Again, the main exception to the rule is the perception of attractiveness in male high ASD scorers - where observers were inaccurate in their categorisation of ASD. The perceptions of attractiveness may link to another popular psychological theory, the halo effect (Dion, 1972), proposing negative personality traits are attributed too less attractive individuals. The interaction between facial symmetry, attractiveness and perceptions of personality traits may go some way towards explaining the ability to accurately detect mental health traits from facial images. However the studies in this thesis all utilised composite stimuli, which by their nature average out individual differences and may have an impact upon facial symmetry. To better understand the links between Extraversion, attractiveness, symmetry and mental health future studies should use individual faces and GMM analysis, although this comes with obvious ethical implications. In addition it must be noted that if the multiple comparisons used to determine perceptions of personality traits are not deemed to be planned contrasts, and corrections such a Holm's sequential Bonferroni (Holm, 1979) are applied a drop in significance is noted across studies, reducing the strength of these findings. This is a flaw in doing exploratory work with both genders and across all five personality traits as well as attractiveness, increasing the number of comparisons made within each study. Future work could concentrate on the personality traits highlighted within these studies to be consistently attributed to individuals with mental health disorders, such as Extraversion, and disregard traits such as Conscientiousness which has yielded few significant results.

Perceptions of gender in mental health disorders

In chapters seven and eight a more novel spin on the previous chapters is taken, determining if there are further perceptions of individuals with mental health disorders, and whether these hold a clue to an underlying deficit across different psychiatric disorders. Firstly, I found support for Baron-Cohen's masculinity hypothesis of ASD, in that males with high AQ scores were perceived to be more masculine in appearance, and in chapter eight I extended this to take into account Badcock and Crespi's diametric hypothesis, demonstrating a perception of femininity in a number of 'psychotic' disorders.

Although separate theories, the extreme male brain and diametric theories are not distinct in concept. The extreme male brain theory can be thought of as the 'mechanistic' end of the diametric theories mechanistic to mentalistic continuum, a concept my findings lend some support to. The findings of chapter eight highlight a distinction between perceptions of ‘mentalistic’ or ‘feminine’ disorders such as depression and eating disorders and the ‘mechanistic’ or ‘masculine’ disorder of ASD. However, due to the limited understanding of the long-term effects of oestrogen in males and testosterone in females the results are limited to individual sexes - reducing the generalisability of my findings.

Although these findings do lend support to the concept of a continuum linking these two theories, it is more socially based. Baron-Cohen has amassed a large collection of evidence to suggest that individuals with ASD display behavioural characteristics deemed to be masculine. A line of investigation which could lend further support for the collaboration of the two theories would be to determine if females with ‘psychotic’ disorders also exhibit behavioural traits considered to be more ‘feminine’. Initial evidence has surrounded the investigation of empathy, with mixed results, but further studies such as determining if women with psychotic disorders are more sensitive to facial expressions or are better sharers - behavioural traits proposed to be feminine (Baron-Cohen, 2002) could provide more evidence.

The results of chapter eight suggest that individuals with ‘extreme female brain’ disorders are perceived in a different way to those with ‘extreme male brain’ disorders, what effect could this have on symptom development and continuation?

Consequences of perceptions of mental health - What came first?

As discussed, there appears to be a distinct perception of individuals with mental health disorders. Not only are observers able to accurately discriminate between high and low scorers on a range of mental health traits, they also consistently attribute socially desirable personality traits to high scorers and alter perceptions of gender. The question to answer here is, what are social the implications of these perceptions?

Perceptions of mental health: Firstly, the ability to detect mental health traits from static facial images has, in itself, social implications, as it suggests the common misconception that mental health disorders are 'hidden' disorders is not entirely true. The ability to detect mental health traits has implications for social stigma, in that individuals with such traits are often perceived less favourably than those with (Canu, Newman, Morrow, & Pope, 2008). However, it is important to note that in this thesis I am not suggesting that by simply viewing an individual's face someone can accurately diagnose them with a specific mental health disorder. These studies merely highlight an ability to discriminate between individuals at either ends of a symptom spectrum, and although accuracy rates are above chance, generally ranging from 55 - 70%, they are still far from perfect. There is an interesting question to ask here: which comes first, do individuals at risk of developing mental health disorders exhibit observable traits which alters how others socially interact with them, or do they have specific traits which effect social interaction and lead to a vulnerability to mental health problems?

Perceptions of personality: High symptom scorers were generally perceived to be lower in Extraversion and Agreeableness, whilst higher in Neuroticism. These are all traits which have, over multiple studies, been implicated in less successful social relationships. These effects go back as far as primary school, for example, Jensen-Campbell et al. (2002) found Agreeableness and Extraversion to be positively related to peer acceptance and friendship. Following through to high school, Van der Linden, Scholte, Cillessen, Nikenhuis and Segers (2010) demonstrated a positive correlation between extraversion and likeability and popularity. They also affect adult life, with higher scorers of Neuroticism reporting lower pay and fewer opportunities for promotion (Judge, Higgins, Thorsen, & Barrick, 1999). A range of studies have also demonstrated a link between low Extraversion, high Neuroticism and reduced social success, such as Lopes, Salovey and Straus (2003) finding they highly correlated with reduced satisfaction with interpersonal relationships.

These perceptions do not just affect social situations, they can have long-lasting impacts; for example preadolescents who experienced higher levels of peer rejection are more likely to experience psychopathological symptoms in adulthood (Bagwell, Newcomb, & Bukowski, 1998).

Perceptions of gender: Perceptions of gender can also have an impact on how individuals are socially received. For males, ratings of facial masculinity have been found to be positively associated with perceptions of aggression (Carre, McCormick, & Mondloch, 2009) and dominance (Perrett et al., 1998), and negatively associated with warmth, emotionality, honesty, cooperativeness and parental quality (Perrett et al., 1998). Females with a feminine facial appearance are seen to be more attractive and increasing masculinity in female faces reduces perceptions of warmth, emotionality, honest and cooperativeness (Perrett et al., 1998).

It is hard to determine which of these perceptions came first, and which play a critical role in the development of mental health disorder symptoms. What is clear is that observers are readily able to detect traits from facial images, and that each of these perceptions can have a notable impact on social interaction. This could cause a circle, where unwittingly signaling symptoms of mental health disorders results in social difficulties, which exasperate the symptoms of these disorders. This thesis highlights that individuals with mental health disorders are perceived in a negative light compared to those without symptoms of mental health. Although the consequences and real life applications of these findings are still unknown the knowledge may help aid interventions to reduce its effects.

Why has this ability developed? Links to wider evolutionary theory

Many psychiatric disorders have a heritability component, leading to a widely asked question, if psychiatric disorders have a negative impact on individual's lives, why have they not been selected out through generations? One evolutionary theory of the origin of psychiatric disorders, the infection-disease hypothesis of depression, suggests that depression has evolved as a behavioural response to pathogens (Kinney & Tanaka, 2009). In a recent review Ander, Tanaka and Kinney (2013) present evidence for the advantageous use of moods, and their ability to create an array of physical and behavioural responses, as an adaptive response, aiding individuals to both fight existing infections and avoid new ones. Evidence is drawn from a range of conditions associated with depression, such a seasonal changes, hormone fluctuations and chronic diseases to support the conclusion that depressive symptoms are an evolved mechanism, aiding the sufferer to withdraw from social life when there is an immune threat, with advantages in both reducing the chance of catching illnesses, and also reducing the threat of spreading these illnesses to kin. Indeed a recent study has suggested that mental illness is implicitly associated with disease, and that this association is exacerbated when individuals have had a recent illness (Lund & Boggero, 2014).

If it is the case that psychiatric disorders have developed as a behavioural response to pathogen presence, why have we evolved to be able to discriminate between individuals with and without the disorders? Or is the ability a reflection or consequence of a more general aspect of mate selection?

There are a number of papers which have suggested the ability to detect physical health from static facial images (Little, McPherson, Dennington, & Jones, 2011), and additional papers suggesting that variables such as the foods we eat can affect skin colouration, which has a knock on effect of perceptions of health (Stephen, Coetzee, & Perrett, 2011). Indeed, if the infection-disease hypothesis is correct, then perceptions of mental-health should logically also reflect perceptions of physical health.

If we are to take into account the associations between genetic fitness and physical health, and combine with the proposals of the infection-disease hypothesis, it would suggest that the ability to detect cues of mental health in faces would provide an evolutionary advantage for the receiver. It is plausable to suggest that these inferences may have developed from the accurate perception of physical health, as individuals with poor physical health are likely to have become a burden on group resources. Indeed if individuals with mental health issues were more susceptible to pathogens then developing a negative perception of them, in the form of attributing negative personality traits, could be advantageous, even if there is a degree of truth within the perceptions. It is interesting to associate this theory with the actual personality traits observers attributed to faces of individuals with depressive symptoms, namely the perceptions of Extraversion. If we are to believe the infection-disease hypothesis then it would be logical to suggest the avoidance behaviour associated with pathogen avoidance and social withdrawal to be likened with Introversion. This has even more power when combined with the findings in chapter six, that observers misattributed perceptions of ASD in males as well as judgement them highly in Extraversion. Indeed there is a vast array of evidence suggesting ASD is a downstream behavioural consequence of extreme testosterone levels, which as discussed earlier may be an honest signal of genetic fitness.

Combining all these theories and evidence together it could be suggested that the ability to detect mental health in faces has developed as an evolutionary adaptive strategy to reduce the potential negative impact on group survival. If individuals with symptoms of mental health disorders are less genetically fit, and exhibit behavioural symptoms as pathogen avoidance, then developing a sensitivity to cues of these traits could be advantageous.

Connectivity between results

The current thesis bases the social perceptions of individuals with mental health traits on judgments of the big five personality traits. However, previous research by Oosterof and Todorov (2008) identified that perceptions of traits from facial images could be summarised by two orthogonal concepts, trustworthiness and dominance. Indeed, additional work by Frume (2012) demonstrated that ratings of trustworthiness and dominance could significantly predict ratings of criminality, and furthermore that manipulating the valence (trustworthiness) of faces had an impact on their ratings of criminality, with angry faces being rated as more criminal than happy faces. Could it be possible that the results found within this thesis can be attributed to this simple theory? Suggesting that observers are not able to uniquely identify each mental health disorder, but instead use a scale of social desirability to infer traits?

As discussed, across chapters two to six there is a consistent pattern of negative personality attributions to high scoring images where accuracy is also found in the mental health trait. That is, when observers are accurate in their discriminations between high and low composite images they commonly also attributed negative personality traits to the high scorer. This suggests that there is a commonality between the attributions of each trait, with high scorers being perceived to be lower in social desirability. This is increasingly apparent in the case of males with ASD, where observers attribute positive personality traits to high scorers and are also inaccurate in their discrimination - misattributing high ASD symptoms to low scorers.

The concept of a continuum underlying the perceptions of the disorders studied in this thesis is not new, indeed chapter eight assesses Badcock and Crespi's (2008) proposal that mental health can be thought of as a continuum ranging from extreme masculinity to extreme femininity. Here it was shown that males with ASD were perceived to be more masculine, where as females with psychotic disorders such as depression, eating disorders and SPD were perceived as more feminine. It is interesting to note that Oosterof and Todorov (2008) proposed masculinity to be a key cue to perceptions of dominance. Which becomes increasingly relevant when, once again, looking at the males with ASD, where observers rated high ASD scorers as more masculine and lower in ASD symptoms than low scorers.

These findings suggest the possibility of a continuum underlying perceptions of mental health, which can be encompassed with both Badcock and Crespi's theory of a gender continuum and associated with Oosterof and Todorov's two-dimensional account of facial attributions. Indeed, as shown within Appendix B, there are high correlations between individuals scoring highly on each measure of mental health and a degree of overlap between the individual photographs used to create each of the composite images. This suggests that there is likely to be a common theme underlying both individuals with mental health problems (possibly explained by Badcock and Crespi's theory) and how they are perceived by naive observers (possibly explained by Oosterof and Todorov's two-dimensional hypothesis).

This does not take away from the social implications of the findings of this thesis, in that the common theme across studies suggests that individuals perceived to be high in mental health traits are also deemed to be low in desirable personality traits. Regardless of the specificity of ratings, the general downgrading of personality traits attributed to individuals with symptoms of mental health could have social implications and further impact upon an individuals symptoms. However, further research should be conducted to determine if the ability to detect mental health traits is unique to each trait, or if observers are using a generalised cue, which although within this thesis appears to show a specific ability, is actually a reflection of a general perception. This could be addressed by presenting observers with high and low composites from a variety of traits simultaneously and asking them to pick the image high in a specific trait. If the ability is a reflection of a general perception of social desirability, or a combination of trustworthiness and dominance, then it would be expected that observers would be accurate in discriminating between high and low composites but unable to accurately identify the individual image corresponding to a specific disorder. However, if the ability is a true reflection of an ability to utilise cues related to specific disorders then observers would be entirely accurate in their choices. In addition it would be enlightening to gain Likert ratings of trustworthiness and dominance alongside perceptions of trait occurrence in a similar style to Frome (2012), this methodology would allow for a regression analysis to determine if perceptions of mental health can be predicted from ratings of dominance and trustworthiness, conclusions that cannot be drawn from the methodologies utilised in this thesis.

Further research

This thesis focused on investigating the ability to detect mental health and the implications this had for perceptions of personality traits across a range of disorders. This approach was undertaken to shed light on the possibility that individuals with a range of psychiatric disorders may face social stigma due to how observers perceive them in the first instance. However, it has left a gap for further research to drill down and identify the specific cues utilised by observers to make these inferences. Geometric morphometric (GMM) analysis could be used to detect differences in facial structure between individuals, and groups. The application of GMM analysis with the samples used in this thesis could help to shed light on the particular aspects of the facial images that are being utilised in the attribution of mental health symptoms. Not only could GMM be used to determine if there is a statistically significant difference in face shape between high and low scorers within each trait, but it could also be used to determine if there is a general face shape that is associated with good and poor general mental health. If results were found to be non-significant , this would suggest that facial colour and texture are an important cue to these symptoms, another viable avenue of research as they have been demonstrated to be important in the accurate identification of personality traits (Jones, Kramer, & Ward, 2012).

GMM could also be used to confirm or disprove some of the concepts discussed as potential explanations for the ability to accurately identify symptoms. Firstly, the analysis could also produce an estimation of symmetry to determine if facial asymmetry is correlated with mental health symptoms in general, and by specific disorders. This could open up answers as to how perceptions of Extraversion, symptoms of mental health and symmetry are intertwined. In addition this methodology would overcome the interference composite images have on symmetry, and as the use of GMM analysis does not require observers to view the stimuli, so individual faces could be used without ethical infringement.

Secondly, GMM could be used to develop further evidence for the gender theories of mental health, such as the diametric hypothesis. Previous studies have utilised algorithms to determine the degree of masculinity within a face which could be used to conclude if males with ASD have a more masculine and females with depression and eating disorders a more feminine face shape. This could also help to shed light on the differing roles sex hormones play in these disorders and lead investigations into the possibility of an underlying continuum across all mental health disorders.

In previous research trained psychotherapists were found to be no more accurate than naïve observers in the identification of individuals who would later go on to commit suicide (Kleinman & Rule, 2013). It would be interesting to determine if clinicians, such as clinical psychologists and psychiatrists, would show improved accuracy in the detection of common disorders such as those in this thesis, shedding light on whether individuals can develop the ability to recognise facial cues through increased interaction with individuals and a specialist knowledge of the disorders. Additional research could investigate whether these 'experts' would also attribute the negative social traits seen to be inferred by naïve observers, or if they are utilising other cues or possibly have a reduced social stigma towards these individuals. Although this research is in its infancy, with a deeper understanding as to what is driving the cues to these disorders could be used in the future to aid the early recognition of individuals with higher likelihood of developing psychiatric disorders.

This thesis has highlighted the potential social implications of the negative personality traits attributed to individuals with symptoms of psychiatric disorders, but could this knowledge be utilised to develop an intervention? Indeed previous research has found that behavioural training in the perception of facial emotions increased positive affect in students scoring highly for depression (Penton-Voak, Bate, Lewis, & Munafo, 2012). Interventions could use a mixture of education into the negative stereotype associated with psychiatric disorders, and the first impressions made of those with symptoms as well as a behavioural mechanism aiming to cut through the negative cycle of depressive symptoms and social rejection.

A final thread of potential future research would be to follow individuals over a series of months, collecting information on symptoms and photographs. This could give rise to a wealth of potential studies, assessing if observers are able to detect which individuals will have increased or decreased symptom scores over time.

Limitations

One key limitation of this thesis is the consistent use of a student participant sample. Although across all chapters there are individuals who score well within the clinical range for each disorder, this is determined using self-report questionnaires and not via clinical assessment. Whilst there are obvious limitations to using a non-clinical sample, in this thesis it has allowed for the collection of a large photo database (exceeding 700 individuals) within a relatively short time span. As preliminary studies they yield interesting results which are worthy of clinical replication, but the generalisability of the current thesis findings must be interpreted with caution.

Another potential limitation to studies within this thesis is the lack of control for postural cues, a variable which has been shown to influence perceptions of personality traits. For example Jones et al. (2012) found perceptions of Extraversion to be reduced to chance level when postural cues, such as slight head tilts, in 3D faces were controlled for.

A final limitation to this study concerns the ethological validity of the methodology. Throughout all chapters composite facial images were used as stimuli, and although this was done to protect anonymity and meet ethical requirements, it does not present stimuli that we would encounter in every day life. Scott and Penton-Voak (2011) highlight this issue when they demonstrated a stronger preference for masculinity when raters viewed composite images, findings with particular implications for results of chapter 7. In addition the majority of studies within this thesis have utilised two-alternative-force-choice methodology, where observers are presented with a pair of faces, one representing individuals with high symptom presence, the other low. This methodology has come under focus recently in a paper by Todorov and Porter (2014), who demonstrated levels of within person variance (the variance found when different observers rate the same face) often reached the same level as between person variance (the variance found when averaging results from a set of stimuli). These findings suggest that it may not be cues from individual faces which drive the apparent ability to recognise traits from faces, but the comparison with another face. A possible way of shedding light on this would be to ask observers to rate individual composite images for symptom severity, and determine if there is a significant difference in judgements from the high and low symptom scorer groups.

Implications

As I have already discussed the findings of this thesis best lend themselves to developing an understanding into the social consequences of perceptions of mental health. The findings that individuals with symptoms of mental health disorders are seen to be unintentionally signaling socially undesirable personality traits, combined with previous research detailing the social implications of these perceptions, suggests that they may enter a repetitive cycle of symptoms and social problems. It would be interesting to determine if these negative social perceptions were upheld even when individuals are a-symptomatic, to determine if this is a temporary consequence or a long-term issue.

Developing an understanding of the social consequences of these perceptions could help to develop interventions for individuals with mental health disorders, enabling them to better understand the social problems they are facing. Additionally increasing the awareness of the general public to the existence of this phenomenon could promote better understanding of the difficulties faced by individuals with mental health problems, and possibly reduce the effect of quick judgements.

Another potential implication of this research comes from a clinical diagnostic angle. If GMM analysis yields a positive result, suggesting there are specific facial features associated with either mental health problems as a whole, or specific disorders, then this could be used to aid early recognition of high-risk individuals. That is not to say that all individuals with these traits will go on to develop mental health problems, but early identification of high risk individuals could reduce waiting times between symptom onset and seeking treatment.

Conclusions

We encounter faces in everyday social interactions and have evolved to make judgments based on this basic stimuli within the first 100ms of interaction (Willis & Todorov, 2006). This thesis extends previous research to demonstrate that naïve observers are able to accurately categorise high and low scorers of a wide range of common mental health disorders from static facial images, and in the case of depressive symptoms this ability is still apparent when the stimuli is greatly impoverished - showing just the eye and brow region. In addition to this a negative social trait attribution is found for most disorders, suggesting that even from first glace, individuals with mental health problems are at a social disadvantage. In addition I assessed how these findings can be entwined with existing theories of the origins of psychiatric disorders to shed light on a potential continuum underpinning psychiatric disorders. Further research should utilise GMM analysis to determine if these perceptions are based on a physical shape difference, and develop predictions into the specific facial features used in this ability.

References

Abbas, Z., & Duchaine, B. (2008). The role of holistic processing in judgments of facial attractiveness. *Perception, 37,* 1187-1196.

Aldridge, K., George, I. D., Cole, K. K., Austin, J. R., Takahashi, T. N., . . . Miles, J. H. (2011). Facial phenotypes in subgroups of pre-pubertal boys with autism spectrum disorders are correlated with clinical phenotypes. *Molecular Autism, 2*(1)*,* 15.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief “red flags” for autism screening: The short autism spectrum quotient and the short quantative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(2), 202-212.

Ambady, N. (2010). The perils of pondering: Intuition and thin slice judgements. *Psychological Inquiry, 21*(4), 271-278.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgements of sexual orientation from thin slices of behaviour. *Journal of Personality and Social Psychology, 77*(3), 538-547.

American Psychiatric Association. (1994). Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Author, Washington DC.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Author, Washington DC.

Anders, S., Tanaka, M., & Kinney, D. K. (2013). Depression as an evolutionary strategy for deference against infection. *Brain, Behavior, and Immunity, 31*, 9-22.

Arato, M., Frecska, E., Beck, C., An, M., & Kiss, H. (2004). Digit length pattern in schizophrenia suggests disturbed prenatal hemispheric lateralization. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 28*(1), 191-194

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258-290

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Difference 38*(2), 451-460.

Autism Education Trust (2015, August 8). What is Autism? Retrieved from http://www.autismeducationtrust.org.uk/about-aet/what-is-autism.aspx

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., & Hackett, G. (2009). Fetal testosterone and autistic traits. *British Journal of Psychology, 100,* 1-22.

Badcock, C. (2009). The imprinted brain: How genes set the balance between autism and psychosis. Jessica Kingsley Publishers, London.

Badcock, C., & Crespi, B. (2006). Imbalanced genomic imprinting in brain development: An evolutionary basis for the aetiology of autism. *Journal of Evolutionary Biology, 19*(4), 1007-1032.

Badcock, C., & Crespi, B. (2008). Battle of Sexes may set the brain. *Nature, 454*(7208), 1054-1055.

Bagby, R. M., Bindseil, K. D., Schuller, D. R., Rector, N. A., Young, T., … Joffe, R. T., (1997). Relationship between the five-factor model of personality and unipolar, bipolar and schizophrenic patients. Psychiatry Research, *70*(2), 83-94.

Bagby, R. M., Costa, P. T., Widiger, T. A., Ryder, A. G., & Margarita, M. (2005). DSM-IV personality disorders and the five-factor model of personality: A multi-method examination of domain-and facet-level predictions. *European Journal of Personality, 19*(4), 307-324.

Bagby, R. M., Young, L. T., Schuller, D. R., Bindscil, K. D., Cooke, R. G., & Dickens, S. E. (1996). Bipolar disorder, unipolar depression and the five-factor model of personality. *Journal of Affective Disorders, 41,* 25-32.

Bagwell, C. L., Newcomb, A. F., & Bukowski, W. M. (1998). Preadolescent friendship and peer rejection as predictors of adult adjustment. *Child Development, 69*(1), 140-153.

Bailey, A. A., & Hurd, P. L. (2005). Depression in men is associated with more feminine finger length ratios. *Personality and Individual Differences, 39*(4), 829-836.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., … Rutter, M. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine, 25*, 63-77.

Baird, G. B., Simonoff, E., Pickles, A., Chandler, S., Locas, T., … Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: The special needs and autism project (SNAP). *The Lancet, 268*(9531), 210-215.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Science, 6*, 248-254.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high – functioning autism, and normal sex differences. *Philosophical transactions of the Royal Society of Biological Sciences, 358,* 361-374.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163- 175.

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

Battaglia, M., & Torgersen, S. (1996). Schizotypal disorder: At the crossroads of genetics and nosology. *Acta Psychiatrica Scandinavica, 94*(5), 303-310.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II Manual*. Psychological Corporation, San Antonia, Texas.

Bedwell, J. S., & Donnelly, R. S. (2005). Schizotypal personality disorder or prodromal symptoms of schizophrenia? *Schizophrenia Research, 80*(2-3), 263-269.

Bejerot, S., Eriksson, J. M., Bonde, S., Carlstrom, K., Humble, M. B., & Eriksson, E. (2012). The extreme male brain revisited: gender coherence in adults with autism spectrum disorder. *The British Journal of Psychiatry, 201,* 116-123.

Bery, D. S., & McArthur, L. Z. (1985). Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48*(2), 312-323.

Berry, D. S., & Wero, J. L. F. (1993). Accuracy in face perception: A view from ecological psychology. *Journal of Personality, 61*(4), 497-520.

Berry, D. S., Willingham, J. K., & Thayer, C. A. (2000). Affect and personality as predictors of conflict and closeness in young adults' friendships. *Journal of Research in Personality, 34*(1), 84-107.

Boggs, C. D., Morey, L. C., Skodol, A. E., Shea, M. T., Sainslow, C. A. … Gunderson, J. G. (2009). Differential impairment as an indicator of sex bias in DSM-IV criteria for four personality disorders *Personality Disorders: Theory, Research, and Treatment, S(*1), 61-68.

Boothroyd, L. G., Jones, B. C., Burt, D. M., DeBruine, L. M., & Perrett, D. I. (2008). Facial correlates of sociosexuality. *Evolution and Human Behavior, 29*(3), 211-218.

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings and measured intelligence. *Journal of Personality and Social Psychology, 65*(3), 546-553.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*(4), 599-614.

Bresmer, J. A., & Gallup, G. G. (2012). From extreme to the other: Negative evaluation anxiety and disordered eating as candidates for the extreme female brain. *Evolutionary Psychology, 10*(3), 457-486.

Brosnan, M., Daggar, R., & Collomosse, J. (2010). The relationship between systemising and mental rotation and the implications for the extreme male brain theory of autism. *Journal of Autism and Developmental Disorders, 40*(1), 1-7.

Brugha, T. S., McManus, S., Bankart, J., Scott, F., Purdon, S., Smith, J., . . . Meltzer, H. (2011). Epidemiology of autism spectrum disorders in adults in the community in England. *Archives of General Psychiatry, 68*, 459-466.

Bunevicius, A., Katkute, A., & Bunevicius, R. (2008). Symptoms of anxiety and depression in medical students and in humanities students: Relationship with big-five personality dimensions and vulnerability to stress. *International Journal of Social Psychiatry, 54*(6), 494-501.

Canu, W. H., Newman, M. L., Morrow, T. L., & Pope, D. L. W. (2008). Social appraisal of adult ADHD: Stigma and influences of the beholder's big five personality traits. *Journal of Attention Disorders, 11*(6), 700-710.

Carney, D. R., Colvin, R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*(5), 1054-1072.

Carre, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behaviour. *Psychological Science, 20*(10), 1194 - 1198.

Chowen, J. A., Argenta, J., Gonzalez-Parra, S., & Garcia- Segura, L. M. (1993). Differential effects of the neonatal and adult sex steroid environments on the organization and activation of hypothalamic growth hormone-releasing hormone and somatostatin neurons. *Endocrinology, 133*, 2792- 2802.

Cosgrove, L., & Riddle, B. (2004). *Gender bias and sex distribution of mental disorders in DSM-IV-TR.* In Caplan, P. J., & Cosgrove, L (Eds.), *Bias in psychiatric diagnosis* (pp. 127-140). New York: The Rowman & Littlefield publishing group.

Costa, P. T., & McCrae, R. R. (1990). Personality disorders and the five-factor model of personality. *Journal of Personality Disorders, 4*(4), 362-371.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO personality inventory. *Psychological Assessment, 4*(1), 5-13.

Costa, P. T., & McCrae, R. R. (1995) Personality traits of Eysenck's P-E-N system: Three- and five-factor solutions. *Journal of Personality and Social Psychology, 69*(2), 308-317.

Crespi, B., & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *The Behavioral and Brain Sciences, 31*(3), 241-261.

Crespi, B., Stead, P., & Elliot, M. (2010). Comparative genomics of autism and schizophrenia. *PNAS, 107*(1), 1736-1741.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97,* 143-156.

Dinsdale, N. (2008). Empathy, Autism and the Extreme Female Brain (MSc dissertation). Simon Fraser University, Canada.

Dinsdale, N., & Crespi, B. J. (2013). The borderline empathy paradox: Evidence and conceptual models for empathic enhancements in borderline personality disorder. *Journal of Personality Disorders, 27*(2), 172-195.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology. 24*(3), 285-290.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment, 18*(2), 192-203.

Dyce, J. A. (1997). The big five factors of personality and their relationship to personality disorders. *Journal of Clinical Psychology, 53*(6), 587-593.

Ekman, P. (1985). *Telling lies: Clue to deceit in the marketplace, politics, and marriage*. WW - Norton & Company, New York.

Endler, N. S., Macrodimitris, S. D., & Kocovski, N. L. (2000). Depression: The complexity of self-report measures. *Journal of Applied Biobehavioural research, 5*(1), 26-46.

Evardone, M., Alexander, G. M., & Morey, L. C. (2008). Hormones and borderline personality features. *Personality and Individual Differences, 44*(1), 278-287.

Fink, B., Neave, N., Manning, J. T., & Grammer, K. (2005). Facial symmetry and the 'big-five' personality factors. *Personality and Individual Difference, 39*(3), 523-529.

Fink, B., Neave, N., & Seydel, H. (2007). Male facial appearance signals physical strength to women. *American Journal of Human Biology, 19*(1), 82-87.

Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science, 11*(5), 154-158.

Fisk, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77-83.

Fisk, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (Dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues, 55*(3), 473-489.

Floderus-Myrhed, B., Pedersen, N., & Rasmuson, I. (1980). Assessment of heritability for personality, based on a short-form of the Eysenck personality inventory: A study of 12,898 twin pairs. *Behavior Genetics, 10*(2), 153-162.

Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PLoS One, 7*(6), e37253.

Fombonne, E. (1999). The epidemiology of autism: A review. *Psychological Medicine, 29*, 769-786.

Fossati, A., Raine, A., Carretta, I., Leonardi, B., & Maffei, C. (2003). The three-factor model of schizotypal personality: Invariance across age and gender. *Personality and Individual Differences, 35*(5), 1007-1019.

Foster-Green, M., Satz, P., Galer, D. J., Ganzell, S., & Kharabi, F. (1989). Minor physical anomalies in schizophrenia. *Schizophrenia Bulletin, 15*(1), 91-99.

Fowler, K. A., Lilienfeld, S. O., & Patrick, C. J. (2009). Detecting psychopathy from thin slices of behavior. *Psychological Assessment, 21*(1), 68-78.

Friedman, J. N. W., Oltmanns, T. F., & Turkheimer, E. (2007). Interpersonal perception and personality disorders: Utilization of a thin slice approach. *Journal of Research in Personality, 41*(3), 667-688.

Gangestad, S. W., Thornhill, R., & Yeo, R. A. (1994). Facial attractiveness, developmental stability, and fluctuating asymmetry. *Ethology and Sociobiology, 15*(2), 73-85.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The eating attitudes test: Psychometric features and clinical correlates. *Psychological Medicine, 12*(4), 871-878.

Geerts, E., & Brune, M. (2011). On the role of ethology in clinical psychiatry: What do ontogenetic and causal factors tell us about ultimate explanations of depression? In Adriaens, P. R., & De Block, A. (Eds.), *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory* (117-141). New York: Oxford University Press.

Goldberg, L. R. (1992). The development of markers of the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.

Goldberg. L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde., I Deary., F. De Fruyt., & F. Ostendorf (Eds.), *Personality psychology in Europe (Vol. 7, pp 7-28).* Tilburg, The Netherlands: Tilburg University Press.

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Paediatrics, 56*(4), 544-549.

Grady, K. L., Phoenix, C. H., & Young, W. C. (1965). Role of the developing rat testis in differentiation of the neural tissues mediating mating behaviour. *Journal of Comparative and Physiological Psychology, 59*(2), 176-182.

Grossman, R. B. (2014). Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. *Autism.* Advanced online publication, June 12th 2014.

Gurrera, R. J., Dicket, C. C., Niznikiewicz, M. A., Voglmaier, M. M., Shenton, M. E., & McCarley, R. W. (2005). The five-factor model in schizotypal personality disorder, *Schizophrenia Research, 80*(2-3), 243-251.

Guy, J. D., Majorski, L. V., Wallace, C. J., & Guy, M. P. (1983). The incidence of minor physical anomalies in adult male schizophrenics. *Schizophrenia Bulletin, 9*(4), 571-582.

Hammond, P., Forster-Gibson, C., Chudler, A. E., Allanson, J. E., Hutton, T. J., Farrell, S. A., . . . Lewis, M.E.S. (2008). Face-brain asymmetry in autism spectrum disorders. *Molecular Psychiatry, 13*(6)*,* 614-623.

Harrigan, J. A., Wilson, K., & Rosenthal, R. (2004). Detecting state and trait anxiety from auditory and visual cues: A meta-analysis. *Personality and Social Psychology Bulletin, 30*(1), 56-66.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology, 78*(5), 837-852.

Heifner, C. (1997). The male experience of depression. *Perspectives in Psychiatric Care, 33*(2), 10-18.

Hennessy, R. J., Lane, A., Kinsella, A., Larkin, C., O'Callaghan. E., & Waddington, J. L. (2004). 3D morphometrics of craniofacial dysmorphology reveals sex-specific asymmetries in schizophrenia. *Schizophrenia Research, 67*(2-3), 261-268.

Hirschfeld, R. M. A., Williams, J. B. W., Spitzer, R. L., Calabrese, J. R., Flynn, L., … Zajecka, J. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The mood disorder questionnaire. *The American Journal of Psychiatry, 157*(11), 1873-1875.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Holtzman, N. S. (2011). Facing a psychopath: Detecting the dark triad from emotionally-neutral faces, using prototypes from the personality faceaurus. *Journal of Research in Personality, 45*(6), 648-654.

Hönekopp, J., Bartholdt, L., Beier, L., & Liebert, A. (2007). Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: New data and a meta-analytic review. *Psychoneuroendocrinology, 32,* 313-321.

Hyler, S. E., Skodol, A. E., Oldham, J. M., Kellman, H. D., & Doidge, N. (1992). Validity of the personality diagnostic questionnaire-revised: A replication in an outpatient sample. *Comprehensive Psychiatry, 33,* 73-77.

Ingudomnukul, E., Baron-Cohen, S., Wheelwright, S., & Knickmeyer, R. (2007). Elevated rates of testosterone-related disorders in women with autism spectrum conditions. *Hormones and Behaviour, 51*, 597-604.

Jensen-Campbell, L. A., Adams, R., Perry, D. G., Workman, K. A., Furdella, J. Q., & Egan, S. K. (2002). Agreeableness, Extraversion, and peer relations in early adolescence: Winning friends and deflecting aggression. *Journal of Research in Personality, 36*(3), 224-251.

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: The contributions of facial shape, skin texture, and viewing angle. *Journal of experimental psychology: Human Perception and Performance, 38*(6), 1353-1361.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick. M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*(3), 621-652.

Judd, C. M., James-Hawkins, L. J., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgement: Understanding the relations between judgements of competence and warmth. *Journal of Personality and Social Psychology, 89*(6), 899-913.

Kendler, K. S., Myers, J., & Reichborn-Kjennerud, T. (2011). Borderline personality disorder traits and their relationship with dimensions of normative personality: A web-based cohort and twin study. *Acta Psychiatr Scandinavia, 123*(5), 349-359.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., … Walters, E. E. (2005). The World Health Organization adult ADHD self-report scales (ASRS): a short screening scale for use in the general population. *Psychological Medicine, 35*, 245 - 256.

Kinney, D. K., & Tanaka, M. (2009). An evolutionary hypothesis of depression and its symptoms, adaptive value and risk factors. The *Journal of Nervous and Mental Disease, 197*(8), 561-567l.

Kleinman, S., & Rule, N. O. (2013). Detecting suicidality from facial appearance. *Social Psychological and Personality Science, 4,* 453-360.

Klump, K. L., Gobrogge, K. L., Perkins, P. S., Thorne, D., Sisk, C. L., & Breedlove, M. (2005). Preliminary evidence that gonadal hormones organize and activate disordered eating. *Psychological Medicine, 36*(4), 539-546.

Knickmeyer, R., Baron-Cohen, S., Fane, B. A., Wheelwright, S., Mathews, G. A., Conway, G. S., . . . Hines, M. (2006). Androgens and autistic traits: A study of individuals with congenital adrenal hyperplasia. *Hormones and Behaviour, 50,* 148-153.

Kotov, R., Gamex, W., Schmidt, F., & Watson, D. (2010). Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*(5), 768-821.

Kramer, R. S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *The Quarterly Journal of Experimental Psychology, 63*(11), 2273-2287.

Krueger, R. F., Caspi, A., & Moffitt, T. E. (2000). Epidemiological personology: The unifying role of personality in population-based research on problem behaviours. *Journal of Personality, 68(*6), 967-998.

Lake, R. E., Eaves, L. J., Maes, H. H. M., Heath, A. C., & Martin, N. G. (2000). Further evidence against the environmental transmission of individuals differences in neuroticism from a collaborative study of 45, 850 twins and relatives on two continents. *Behavior Genetics, 30*(3), 223-233.

Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemising in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders, 34*, 301-310.

Little, A. C., McPherson, J., Dennington, L., & Jones, B. C. (2011). Accuracy in assessment of self-reported stress and a measure of health from static facial information. *Personality and Individual Differences, 51*(6), 693-698.

Little, A. C., & Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology, 98,* 111-126.

Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality and the perceived quality of social relationships. *Personality and Individual Difference, 35*(3), 641-658.

Lund, E. M., & Boggero, I. A. (2014). Sick in the head? Pathogen concerns bias implicit perceptions of mental illness. *Evolutionary Psychology, 12*(4), 706-718.

Lynam, D. R., & Widiger, T. A. (2001). Using the five-factor model to represent the DSM-IV personality disorders: An expert consensus approach. *Journal of Abnormal Psychology, 110*(3), 401-412.

Malas, M. H., Dogan, S., Evcil, E. H., & Desdicoglu, K. (2006). Fetal development of the hand, digits and digit ratio (2D:4D). *Early Human Development, 82*(7), 469-475.

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment, 27*(2), 101-114.

Manning, J. T., Scutt, D., Wilson, J., & Levis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: A predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Human Reproduction, 13,* 3000-3004.

Mason, A. E., Sbarra, D. A., & Mehl, M. R. (2010). Thin-slicing divorce: Thirty seconds of information predict changes in psychological adjustment over 90 days. *Psychological Science, 21*(10), 1420-1422.

Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine, 3*(11), 2011-2030.

Mareckova, K., Weinbrand, Z., Chakravarty, M. M., Lawrence, C., Aleong, R., Leonard, G., . . . Paus, T. (2011). Testosterone-mediated sex differences in the face shape during adolescence: Subjective impressions and objective features. *Hormones and Behaviour, 60*(5), 681-690.

Maric, N., Krabbendam, L., Vollebergh, W., de Graaf, R., & van Os, J. (2003). Sex differences in symptoms of psychosis in a non-selected, general population sample. *Schizophrenia Research, 63*(1-2), 89-95.

McCrae, R. R., & Costa, P. T. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences, 6*(5), 587-597.

McManus, S., Meltzer, H., Brugha, T., Bebbington, P., & Jenkins, R. (2007). *Adult psychiatric morbidity in England, 2007: Results of a household survey.* The NHS Information Centre for Health and Social Care, Leeds.

Mehl, R. M. (2006). The lay assessment of subclinical depression in daily life. *Psychological Assessment, 18*(3), 340-345.

Meindl, K., Windhager, S., Wallner, B., & Schaefer , K. (2012). Second-to-forth digit ratio and facial shape in boys: The lower the digit ratio the more robust the face. *Proceedings of the Royal Society of Biology, 279*, 2457-2463.

Meuller, U., & Mazur, A. (1997). Facial dominance in Homo sapiens as honest signalling of male quality. *Behavioral Ecology, 8*(5), 569-579.

Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., & Schulte, D. (2009). Embodiment of sadness and depression: Gait patterns associated with dysphoric mood. *Psychomatic medicine, 71*(5), 580-587.

Miles, J. H., & Hillman, R. E. (2000). Value of a clinical morphology examination in Autism. *American Journal of Medical Genetics, 91*(4), 245-255.

Millon, T. (1987). *Manual for the millon clinical multiaxial inventory-II (MSMI-II).* Minneapolis, MN: National Computer Systems.

Moller, A. P., & Swaddle, J. P. (1997). *Asymmetry, developmental stability and evolution.* Oxford: Oxford University Press.

Morey, L. C., Gunderson, J. G., Quigley, B. D., Shea, M. T., Skodol, A. E. … Zanarini, M. C. (2002). The representation of borderline, avoidant, obsessive-compulsive and schizotypal personality disorders by the five-factor model. *Journal of Personality Disorders, 16*(3), 215-234.

Morey, L. C., Waugh, M. H., & Blashfield, R. K. (1985). MMPI scales for DSM-III personality disorders: their derivation and correlates. *Journal of Personality Assessment, 49*(3), 245-251.

Muntaner, C., Garcia-Sevilla, L., Fernandez, A., & Torrubia, R. (1988). Personality dimensions, schizotypal and borderline personality traits and psychosis proneness. *Personality and Individual Differences, 9*(2), 257-268.

Neave, N., Laing, S., Fink, B., & Manning, J. T. (2003). Second to forth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society of Biological Sciences, 270*, 2167 – 2172.

O'Callaghan, E., Buckley, P., Madigan, C., Redmond, O., Stack, J. P., …Waddington, J. L. (1995). The relationship of minor physical anomalies and other putative indices of developmental disturbance in schizophrenia to abnormalities in cerebral structure on magnetic resonance imaging. *Biological Psychiatry, 38*(5), 516-524.

O'Connor, L.E., Berry, J.W., Lewis, T., Mulherin, K., & Yi, E. (2007). Empathy and depression: The moral system on overdrive. In T.F.D. Farrow and P.W.R. Woodruff (Eds.), *Empathy and Mental Illness.* Cambridge: Cambridge University Press.

Oltmanns, T. F., Friedman, J. N. W., Fiedler, E. R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behaviour. *Journal of Research in Personality, 38*(3), 216-229.

Ormel, J., & Wohlfarth, T. (1991). How neuroticism, long-term difficulties, and life situation change influence psychological distress: A longitudinal model. *Journal of Personality and Social Psychology, 60*, 744-755.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *PNAS, 105*(32), 11087-11092.

Ozgen, H., Hellemann, G. S., de Jonge, M. V., Beemer, F. A., & van Engeland, H. (2013). Predictive value of morphological features in patients with autism versus normal controls. *Journal of Autism and Developmental Disorders, 43,* 147-155.

Ozgen, H. M., Hellemann, G. S., Stellato, R. K., Lahuis, B., van Daalen, E., Staal, W. G., . . . van Engeland, H. (2011). Morphological features in children with autism spectrum disorders: A matched case-control study. *Journal of Autism and Developmental Disorders, 41*(1), 23-31.

Ozgen, H. M., Hop, J. W., Hox, J. J., Beemer, F. A., & Engeland, H. (2010). Minor physical anomalies in autism: a meta-analysis. *Molecular Psychiatry, 15*(3), 300- 307.

Paunonen, S. V., Ewan, K., Earthy, J., Lefave, S., & Goldberg, H. (1999). Facial features as personality cues. *Journal of Personality, 67,* 555-583.

Penke, L. (2011). The revised sociosexual orientation inventory. In T.D. Fisher, C. M. Davis, W. L. Yarber & S. L. Davis (Eds): *Handbook of Sexuality related measures* (3rd Ed.). Taylor & Francis: New York.

Penton-Voak, I. S., Bate, H., Lewis, G., & Munafo, M. R. (2012). Effects of emotion perception training on mood in undergraduate students: Randomised controlled trial. *The British Journal of Psychiatry, 201*(1), 71-72.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behaviour, 25*, 229-241.

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a “kernel of truth” in social perception. *Social Cognition, 24,* 607-640.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., … Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394*(6696), 884-887.

Phoenix, C. H., Goy, R. W., Gerall, A. A., & Young, W. C. (1959). Organizing action of prenatally administered testosterone proportionate on the tissues mediating mating behaviour in the female guinea pig. *Endocrinology, 65*(3), 369-382.

Pound, N., Penton-Voak, I. S., & Brown, W. M. (2007). Facial symmetry is positively associated with self-reported extraversion. *Personality and Individual Differences, 43*(6), 1572-1582.

Preti, A., de Girolamo, G., Vilagut, G., Alonso, J., de Graaf, R., … Morosini, P. (2009). The epidemiology of eating disorders in six European countries: Results of the ESEMeD-WMH project. *Journal of Psychiatric research, 43*(14), 1125 -1132.

Putz, D. A., Gaulin, S. J. C., Sporter, R. J., & McBurney, D. H. (2004). Sex hormones and finger length: What does 2D:4D indicate? *Evolution and Human Behaviour, 25*(3), 182-199.

Quist, M. C., Watkins, C. D., Smith, F. G., DeBruine, L. M., & Jones, B. C. (2011). Facial masculinity is a cue to women's dominance. *Personality and Individual Differences, 50*(7), 1089-1093.

Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin, 17*(4), 555-564.

Raine, A. (1992). Sex differences in schizotypal personality in a nonclinical population. *Journal of Abnormal Psychology, 101*(2), 361-364.

Raine, A., & Benishay, D. (2005). The SPQ-B: A brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders, 9*(4), 346-355.

Rauthmann, J. F. (2012). The dark triad and interpersonal perception: Similarities and differences in the social consequences of narcissism, Machiavellianism and psychopathy. *Social, Psychological and Personality Science, 3*(4), 487-496.

Roberts, S. B., & Kendler, K. S. (1999). Neuroticism and self-esteem as indices of the vulnerability to major depression in women. *Psychological Medicine, 29*(5), 1101-1109.

Rodier, P. M., Bryson, S. E., & Welch, J. P. (1997). Minor malformations and physical measurements in Autism: Data from Nova Scotia. *Teratology, 55*(5), 319-325.

Rowland, D. A., & Perrett, D. I. (1995). Manipulating facial appearance through shape and colour. *IEEE Computer Graphics and Applications, 15*(5), 70-76.

Rule, N. O., & Ambady, N. (2008a). Brief exposures: Male sexual orientation is accurately perceived at 50ms. *Journal of Experimental Social Psychology, 44*(4), 1100-1105.

Rule, N, O., & Ambady, N. (2008b). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science, 19*(2), 109-111.

Rule, N. O., & Ambady, N. (2010a). First impressions of the face: Predicting success. *Social and Personality Psychology Compass, 4*(8), 506-516.

Rule, N. O., & Ambady, N. (2010b). Democrats and republicans can be differentiated from their faces. *PLoS One, 5*(1), 1-7.

Rule, N. O., & Ambady, N. (2011). Judgements of power from college yearbook photos and later career success. *Social Psychological and Personality Science, 2*(2), 154-158.

Rule, N. O., Ambady, N., Adams, R. B., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of sexual orientation. *Journal of Personality and Social Psychology, 95*(5), 1019-1028.

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental and Social Psychology, 45*(6), 1245-1251.

Rule, N. O., Garrett, J. V., & Ambady, N. (2010). On the perception of religious group membership from faces. *PLoS One, 5*(12), 1-10.

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine, 26*, 477-486.

Ruta, L., Ingudomnukul, E., Taylor, K., Chakrabarti, B., & Baron-Cohen, S. (2011). Increased serum androstenedione in adults with autism spectrum conditions. *Psychoneuroendocrinology, 36,* 1154-1163

Sailsbury Centre for Mental Health. (2010). *The economic and social costs of mental health problems in 2009/10*. Centre for Mental Health, London.

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review, 28*(8), 1326-1342.

Santos, I. M., & Young, A. W. (2011). Inferring social attributes from different face regions: Evidence for holistic processing. *Quarterly Journal of Experimental Psychology, 64,* 751-766.

Saulsman, L. M., & Page, A. C. (2005). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review, 23*(8), 1055-1085.

Schaefer, K., Fink, B., Mitteroecker, P., Neave, N., & Bookstein, F. L. (2005). Visualizing facial shape regression upon 2nd to 4th digit ratio and testosterone. *Collegium Antropologicum, 29,* 415-419.

Schiffman, J., Ekstrom, M., LaBrie, J., Schulsinger, F., Sorensen, H., & Mednick, S. (2002). Minor physical anomalies in schizophrenia spectrum disorders: A prospective investigation. *The American Journal of Psychiatry, 159*(2), 238-243.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *Yearbook of Physical Anthropology, 44,* 3-24.

Scott, I. M. L., Clarke, A. P., Boothroyd, L. G., & Penton-Voak, I. S. (2012). Do men's faces really signal heritable immunocompetence? *Behavioural Ecology, 24*(3), 579-589.

Scott, N. J., Jones. A. J., Kramer, R. S. S., & Ward, R. (2015). Facial dimorphism in Autistic Quotient Scores. *Clinical Psychological Science, 3*(2), 230-241.

Scott, N. J., Kramer, R. S. S., Jones, A. J., & Ward, R. (2013). Facial cues to depressive symptoms and their associated personality attributions. *Psychiatry Research, 208,* 47-53.

Scott, I. M., & Penton-Voak, I. S. (2011). The validity of composite photographs for assessing masculinity preferences. *Perception, 40*(3), 323-331.

Segrin, C. (1992). Specifying the nature of social skills deficits with associated depression. *Human Communication Research, 19*(1), 89-123.

Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M., & Meeus, W. (2010). Emerging late adolescent friendship networks and big five personality traits: A social network approach. *Journal of Personality, 78*(2), 509-538.

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of The Royal Society of Biological Sciences, 276*(1656), 575-584.

Sen, S., Nesse, R. M., Stoltenberg, S. F., Li, S., Gleiberman, L., … Burmeister, M. (2003). A BDNF coding variant is associated with the NEO personality inventory domain neuroticism, a risk factor for depression. *Neuropsychopharmacology, 28*, 397-401.

Shackelford, T. K., & Larsen, R. J. (1997). Facial asymmetry as an indicator of psychological, emotional, and physiological distress. *Journal of Personality and Social Psychology, 72*(2), 456-466.

Shevlin, M., Walker, S., Davies, M. N. O., Banyard, P., & Lewis, A. C. (2003). Can you judge a book by its cover? Evidence of self-stranger agreement on personality at zero acquaintance. *Personality and Individual Differences, 35*(6), 1373-1383.

Simon, G. E., & VonKorff, M. (1998). Suicide mortality among patients treated for depression in an insured population. *American Journal of Epidemiology, 147*(2), 155-160.

Singleton, N., Bumpstead, R., O'Brien, M., Lee, A., & Meltzer, H. (2001). *Psychiatric morbidity among adults living in private households, 2000.* The Stationary Office, London.

Sisk, C. L., Schulz, K. M., & Zehr, J. L. (2003). Puberty: A finishing school for male social behaviour. *Annals of the New York Academy of Sciences, 1007,* 189-198.

Sivkov, S., Akabaliev, V., Mantarkov, M., Ahmed-Popova, F., & Akabalieva, K. (2013). Discriminating value of total minor physical anomaly score on the Waldrop scale between patients with bipolar I disorder and normal controls. *Psychiatry Research, 210*(2), 451-456.

Skodol, A. E., Gunderson, J. G., Pfohl, B., Widiger, T. A., Livesley, W. J., & Siever, L. J. (2002). The borderline diagnosis I: Psychopathology, co morbidity, and personality structure. *Biological Psychiatry, 51*(12), 936-950.

Slepian, M. L., Bogard, K. R., & Ambady, N. (2014). Thin-Slice judgements in the clinical context. *Annual Review of Clinical Psychology, 10,* 131-153.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). Manual for the state-trait anxiety inventory (self-evaluation questionnaire). Palo Alto, CA: Consulting Psychologists Press.

Soldz, S., Budman, S., Demby, A., & Merry, J. (1993). Representation of personality disorders in circumplex and five-factor space: Explorations with a clinical sample. *Psychological Assessment, 5*(1), 41-52.

Stauder, J. E. A., Cornet, L. J. M., & Ponds, R. W. H. M. (2011). The extreme male brain theory and gender role behavior in persons with an autism spectrum condition. *Research in Autism Spectrum Disorders, 5,* 1209-1214

Stephen, I. D., Coetzee, V., & Perrett, D. I. (2011). Carotenoid and melanin pigment coloration affect perceived human health. *Evolution and Human Behaviour, 31,* 216-227.

Stewart, L. H., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology, 141*(4), 715-727.

Stillman, T. F., Maner, J. K., & Baumeister, R. F. (2010). A thin slice of violence: Distinguishing violent from non-violent offenders at a glance. *Evolution and Human Behavior, 31*(4), 298-303.

Stirrat, M., & Perrett., D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349-354.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *The American Journal of Psychiatry, 157(*10), 1152-1562.

Sunnafrank, M., & Ramirez, A. (2004). At first sight: Persistent relational effects of get-acquainted conversations. *Journal of Social and Personal Relationships, 21*(3), 361-379.

Tanner, J. M. (1989). *Foetus into man: Physical growth from conception to maturity*. Castlemead : Hertforshire, UK.

Teasdale, J. D. (1998). Cognitive vulnerability to persistent depression. *Cognition and Emotion, 2*(3), 247-274.

Teatero, M. L., & Netley, C. (2013). A critical review of the research on the extreme male brain theory and digit ratio (2D:4D). *Journal of Autism and Developmental Disorders.* Advance online publication.

Tenyi, T., Trixler, M., & Csabi, G. (2009). Minor physical anomalies in affective disorders: A review of the literature. *Journal of Affective Disorders, 112*(1-3), 11-18.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Science, 3*(12), 452-460.

Thornton, J., Zehr, J. L., & Loose, M. D. (2009), Effects of prenatal androgens on rhesus monkeys: A model system to explore the organizational hypothesis in primates. *Hormones and Behaviour, 55*(5), 633-644.

Tiddeman, B. P., Burt, D. M., & Perrett, D. I. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications, 21*, 42–45.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cogntiive and Affective Neuroscience, 3*(2), 119-127.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology, 25*(3), 395-410.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623-1626.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*(6), 813-833.

Todorov, A., & Porter, J. M. (2014). Misleading First Impressions: Different for different facial images of the same person. *Psychological Science, 25*(7), 1404-1417.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Science, 12*(12), 455-460.

Trull, T. J. (1992). DSM-III-R personality disorders and the five-factor model of personality: An empirical comparison. *Journal of Abnormal Psychology, 101*(3), 553-560.

Trull, T. J., & Sher, K. J. (1994). Relationship between the five-factor model of personality and axis I disorders in a nonclinical sample. *Journal of Abnormal Psychology, 103*(2), 350-360.

Trull, T. J., & Widiger, T. A. (1997). *Structured interview for the five-factor model of personality (SIFFM): Professional manual.* Odessa, FL: Psychological assessment resources.

Trull, T. J., Widiger, T. A., & Burr, R. (2001). A structured interview for the assessment of the five-factor model of personality: Facet-level relations to the axis II personality disorders. *Journal of Personality, 62*(2), 175- 198.

Trull, T. J., Widiger, T. A., Lynam, D. R., & Costa, P. T. (2003). Borderline personality disorder from the perspective of general personality functioning. *Journal of Abnormal Psychology, 112*(2), 193-202.

Tskhay, K. O., Feriozzo, M. M., & Rule, N. O. (2013). Facial features influence the categorization of female sexual orientation. *Perception, 42,* 1090-1094.

Valla, J. M., Ceci, S. J., & Williams, W. M. (2011). The accuracy of inferences about criminality based on facial appearance. *Journal of Social, Evolutionary, and Cultural Psychology, 5*(1), 66-91.

Van der Linder, D., Scholte, R. H. J., Cillessen, A. H. N., Nijenhuis, J. T., & Segers, E. (2010). Classroom ratings of likeability and popularity are related to the big five and the general factor of personality. *Journal of Research in Personality, 44*(5), 669-672.

Van Os, J., & Jones, P. B. (2001). Neuroticism as a risk factor for schizophrenia. *Psychological Medicine, 31*(6), 1129-1134.

Van Os, J., Park, S. B. G., & Jones, P. B. (2001). Neuroticism, life events and mental health: Evidence for person-environment correlation. *The British Journal of Psychiatry, 178*(S40), s72-s77.

Vazire, S., Naumann, L. P., Rentfrow, P. J., & Gosling, S. D. (2008). Portrait of a narcissist: Manifestations of narcissism in physical appearance. *Journal of Research in Personality, 42*(6), 1439-1447.

Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *European Journal of Orthodontics, 21,* 137-143.

Vermeersch, H., T'Sjoen, G., Kaufman, J. M., & Vincke, J. (2008). The role of testosterone in aggressive and non-aggressive risk-taking in adolescent boys. *Hormones and Behaviour, 53*(3), 463-471.

Voglmaier, M. M., Seidman, L. J., Niznikiewicz, M. A., Dickey, C. C., Shenton, M. E., & McCarley, R. W. (2005). A comparative profile analysis of neuropsychological function in men and women with schizotypal personality disorder. *Schizophrenia Research, 74*(1), 43-49.

Voracek, M. (2008). Digital ratio (2D:4D) as markers for mental disorder: Low (Masculinized) 2D:4D in autism-spectrum disorders, high (Feminized) 2D:4D in schizophrenia-spectrum disorders. *Behavioral and Brain Sciences, 31*(3), 283-294.

Wakabayashi, A., Baron-Cohen, S., & Wheelwright, S. (2006). Are autistic traits an independent personality dimension? A study of the Autism-Spectrum Quotient (AQ) and the NEO-PI-R. *Personality and Individual Differences, 41*(5), 873-883.

Waldrop, M. F., Pedersen, F. A., & Bell, R. Q. (1968). Minor physical anomalies and behaviour in preschool children. *Child Development, 39*(2), 391- 400.

Walker, H. A. (1977). Incidence of minor physical anomaly in Autism. *Journal of Autism and Childhood Schizophrenia, 7*(2), 165-176.

Wallen, K. (2009). The organizational hypothesis: Reflections on the 50th anniversary of the publication of Phoenix, Goy, Gerall and Young (1959). *Hormones and Behaviours, 55*(5), 561- 565.

Ware, J., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220-233.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Waxer, P. (1976). Nonverbal cues for depth of depression: Set versus no set. *Journal of Consulting and Clinical Psychology, 44*(3), 493.

Weinstein, D. D., Diforio, D., Schiffman, J., Walker, E., & Bonsall, R. (1999). Minor physical anomalies, dermatoglyphic asymmetries, and cortisol levels in adolescents with schizotypal personality disorder. *The American Journal of Psychiatry, 56*(4), 617-623.

Weintraub, K. (2011). Autism Counts. *Nature, 479*(7371)*,* 22-24.

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwals, S., …Yeh, E. K. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association, 276(*4), 293-299.

White, G. M. (1980). Conceptual universals in interpersonal language. *American Anthropologist, 82*, 759-781.

Widiger, T. A., & Trull, T. J. (1992). Personality and psychopathology: An application of the five-factor model. *Journal of Personality, 60*, 363-393.

Widiger, T. A., & Weissman, M. M. (1991). Epidemiology of borderline personality disorder. *Psychiatric services, 42*(1), 1015-1021.

Wiggins, J. S., & Pincus, A. L. (1989). Conceptions of personality disorders and dimensions of personality. *Journal of Consulting and Clinical Psychology, 1*(4), 305-316.

Willis, J., & Todorov, A. (2006). First Impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*(7), 592 - 598.

Wojciszke, B. (1994). Multiple meanings of behaviour: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology, 67*(2), 222-232.

Wojiciszke, B., Dowhyluk, M., & Jaworski, M. (1998). Moral competence-related traits: How do they differ? *Polish Psychological Bulletin, 29*(4), 283-294.

Wolf, M., Limberger, M. F., Kleindienst, N., Stiglitz, R., Domsalla, M., … Bohus, M. (2009). Short version of the borderline symptom list (BSL-23): Development and psychometric evaluation. *Psychotherapie, Psychosomatik, Medizinische Psychologie, 59*(8), 321-324.

Ybarra, O., Chan, E., & Park, D. (2001). Young and old adults' concerns about morality and competence. *Motivation and Emotion, 25*(2), 85-100.

Zahn-Wexler, C., Shirtcliff, E. A., & Marceau, K. (2008). Disorders of childhood and adolescence: Gender and psychopathology. *Annual Review of Clinical Psychology, 4,* 275-303.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

Appendix A - Balanced morph design methodology

In a number of studies I have utilised a balanced morph design, this involves the following methodology:

1. Selecting the highest 18 scorers for the trait
2. Randomly assigning numbers 1-18 to each of these participants
3. Creating nine composite images from groups of 4 of these participants using the matrix below
4. Replicating the procedure for the lowest 18 of the trait, and for each sex.

Morph number	1	2	3	4	5	6	7	8	9
Image 1	2	7	17	12	16	8	1	3	9
Image 2	11	6	4	15	14	18	13	5	10
Image 3	1	11	18	6	15	16	8	2	17
Image 4	4	9	12	3	13	5	10	7	14

Appemdix B - Photo shoot methodology

Three photo shoots were conducted to collect photograph stimuli used within this thesis. During each photo shoot participants were asked to have their photograph taken and complete a number of questionnaires, as described below.

Photo shoot A

Participants: 225 (130 female) participants (age M = 21.45, SD = 5.04) were compensated £5 for their participation.

Measures: Participants completed the following questionnaires:

- Mini International Personality Item Pool (Donnellan, Oswald, Baird & Lucas, 2006)
- Autistic Quotient (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001)
- Short-form health survey (Ware, Kosinski & Keller, 1996)
- Beck's depression inventory (Beck, Steer & Brown, 1996)
- Dominance subscale of the international personality item pool (Goldberg, 1999)

Photo shoot B

Participants: 221 (130 female) participants (age M = 21.65, SD = 5.09) were compensated £6 for their participation.

Measures: Participants completed the following questionnaires:

- Mini International Personality Item Pool (Donnellan, Oswald, Baird & Lucas, 2006)
- Dominance subscale of the international personality item pool (Goldberg, 1999)
- Short-form health survey (Ware, Kosinski & Keller, 1996)

- Positive and negative affect schedule (Watson, Clark & Tellegen, 1988)
- Schizotypal personality questionnaire (Raine, 1991)
- Inventory of depressive symptomatology (Rush, Gullion, Basco, Jarrett & Trivedi, 1996)
- Adult ADHD self-report scale (Kessler et al., 2005)
- State-Trait anxiety inventory (Spielberger, Gorsuch & Lushene, 1970)
- Obsessive-Compulsive Inventory (Foa et al., 2002)
- Mood disorder questionnaire (Hirschfeld et al., 2000)

Photo shoot C

Participants: 240 (141 female) participants (age *M* = 21.69, *SD* = 5.19) were compensated with £10 for their participation.

Measures: Participants completed the following questionnaires:

- Mini International Personality Item Pool (Donnellan, Oswald, Baird & Lucas, 2006)
- Flourishing scale (Diener et al., 2010)
- Scale of positive and negative experience (Diener et al., 2010)
- 10-item Autism Spectrum Quotient (Allison, Auyeung & Baron-Cohen, 2012)
- Sociosexual orientation inventory (Penke, 2011)
- Inventory of depressive symptomatology (Rush, Gullion, Basco, Jarrett & Trivedi, 1996)
- Eating attitudes test (Garner, Olmsted, Bohr & Garfinkel, 1982)
- Short-form health survey (Ware, Kosinski & Keller, 1996)

- Schizotypal personality questionnaire - brief (Raine & Benishay, 2005)
- Borderline symptom list (Bohus et al., 2007)
- Mood disorder questionnaire (Hirschfeld et al., 2000)

Composite image creation

The photographs of the highest and lowest scoring individuals on each questionnaire were subsequently used to create composite stimuli in the research chapters. The specific methodology for stimuli creation can be found within each chapter and the participant numbers used is detailed below.

Photo shoot A

Individuals from this photo shoot were used to create stimuli for study 2.1 and 7.1 and all in chapter 6. The details of the participant numbers used in each composite image are detailed below:

High scoring males

- Depression: 13, 51, 52, 65, 69, 70, 82, 98, 107, 166, 167, 186, 215, 216, 220
- ASD: 6, 52, 65, 69, 82, 88, 93, 98, 107, 115, 167, 215, 216, 219, 225
- ASD - Social skills: 13, 51, 52, 61, 82, 93, 107, 135, 158, 167, 179, 204, 215, 216
- ASD - Communication skills: 52, 61, 65, 69, 70, 82, 88, 129, 135, 158, 163, 167, 179, 216, 219

Low scoring males

- Depression: 34, 37, 63, 77, 81, 88, 93, 99, 101, 133, 135, 163, 179, 184, 225
- ASD: 12, 34, 42, 49, 63, 70, 77, 81, 99, 102, 106, 197, 198, 224
- ASD - Social skills: 6, 12, 37, 49, 63, 77, 81, 88, 99, 106, 133,163, 197, 198, 224

- ASD - Communication skills: 6, 12, 13, 34, 37, 77, 81, 93, 99, 102, 106, 133, 184, 198, 204

High scoring females

- Depression: 9, 10, 32, 57, 62, 68, 78, 95, 116, 117, 149, 173, 181, 212, 218
- ASD: 28, 54, 57, 59, 68, 78, 89, 94, 95, 97, 108, 162, 171, 177, 212
- ASD - Social skills: 28, 54, 57, 59, 78, 79, 89, 94, 95, 97, 108, 116, 171, 189, 212
- ASD - Communication skills: 9, 25, 57, 78, 89, 94, 95, 97, 108, 116, 141, 62, 165, 174

Low scoring females

- Depression: 5, 21, 38, 47, 50, 54, 90, 114, 127, 128, 132, 146, 189, 191, 196, 209,
- ASD: 5, 32, 33, 45, 47, 50, 74, 103, 120, 128, 145, 164, 172, 190, 209
- ASD - Social skills: 5, 50, 55, 84, 105, 127, 128, 148, 164, 190, 191, 193, 207, 209, 217,
- ASD - Communication skills: 50, 55, 74, 84, 104, 119, 120, 128, 146, 150, 164, 188, 191, 208, 209

Photo shoot B

Individuals from this photo shoot were used to create stimuli for study 2.2, 4.1, 7.2, 7.3 and 8. The details of the participant numbers used in each composite image are detailed below:

High scoring males

- Depression: 22, 32, 35, 36, 87, 89, 91, 92, 118, 119, 139, 140, 144, 170, 204
- Schizotypal personality disorder: 2, 22, 32, 34, 62, 63, 89, 91, 92, 118, 124, 139, 140, 44, 148, 157, 170, 181,

- ASD: 22, 23, 25, 32, 62, 63, 81, 91, 92, 118, 134, 39, 141, 144, 157, 181, 220
- ASD social skills: 3, 22, 32, 36, 48, 62, 63, 81, 87, 91, 92, 134, 139, 141, 144, 181, 199, 220,
- ASD communication skills: 2, 32, 36, 47, 62, 63, 70, 81, 87, 91, 92, 139, 141, 144, 181, 199, 204, 220,
- ASD imagination: 1, 22, 23, 25, 32, 62, 63, 87, 91, 134, 141, 147, 181, 185, 199, 204, 220
- ASD attention to detail: 1, 22, 23, 25, 27, 62, 70, 88, 91, 98, 118, 139, 147, 157, 170, 198, 213, 220
- ASD attention switching: 22, 32, 34, 48, 63, 81, 87, 92, 118, 119, 134, 139, 140, 141, 144, 157, 181, 220

Low scoring males:

- Depression: 2, 5, 25, 47, 48, 81, 88, 98, 127, 177, 183, 185, 194, 208, 220
- Schizotypal personality disorder: 1, 4, 17, 47, 48, 87, 88, 98, 114, 141, 147, 177 183, 185, 194, 198, 209, 220
- ASD: 4, 5, 17, 28, 35, 70, 89, 95, 98, 114, 124, 127, 175, 177, 183, 185, 194, 198
- ASD social skills: 1, 2, 4, 5, 17, 23, 28, 35, 70, 89, 95, 98, 114, 124, 170, 183, 185, 213
- ASD communication skills: 4, 5, 17, 35, 48, 88, 89, 95, 98, 114, 124, 127, 140, 147, 177, 185, 194, 198,
- ASD imagination: 4, 5, 17, 28, 34, 35, 48, 70, 88, 89, 95, 98, 114, 119, 175, 177, 194, 198
- ASD attention to detail: 4, 32, 48, 56, 63, 81, 87, 92, 95, 114, 119, 127, 134, 141, 177, 183, 185, 194,

- ASD attention switching: 2, 17, 28, 35, 56, 70, 91, 95, 98, 114, 124, 177, 183, 185, 194, 198, 199, 213

High scoring females

- Depression: 9, 12, 16, 20, 33, 61, 68, 85, 93, 102, 108, 133, 137, 153, 172
- Schizotypal personality disorder: 9, 20, 33, 37, 43, 58, 69, 85, 93, 102, 108, 133, 135, 136, 137, 152, 153, 154
- ASD: 10, 24, 27, 30, 23, 85, 93, 101, 102, 111, 126, 133, 136, 137, 152, 154, 179, 196,
- Bipolar disorder: 9, 10, 37, 67, 68, 80, 85, 86, 93, 109, 122, 150, 151, 153, 163, 168, 193, 219,

Low scoring females:

- Depression: 50, 54, 73, 112, 116, 125, 146, 161, 163, 168, 190, 195, 206, 210, 217
- Schizotypal personality disorder: 73, 99, 112, 131, 146, 151, 160, 161, 168, 189, 190, 192, 193, 195, 197, 206, 210, 215
- ASD: 68, 74, 82, 99, 146, 151, 160, 162, 171, 173, 182, 184, 190, 192, 195, 197, 203, 206,
- Bipolar disorder: 30, 50, 54, 72, 112, 115, 126, 131, 138, 149, 162, 169, 174, 197, 202, 203, 210, 218

Photo shoot C

Individuals from this photo shoot were used to create stimuli for study 4.2, 5 and 8. The details of the participant numbers used in each composite image are detailed below:

High scoring males

- Borderline personality disorder: 6, 22, 24, 28, 37, 49, 52, 53, 60, 88, 122, 159, 164, 188, 196, 197, 218, 254
- Schizotypal personality disorder: 22, 24, 28, 37, 40, 47, 50, 59, 65, 88, 105, 159, 164, 175, 197, 218, 231, 254

Low scoring males

- Borderline personality disorder: 5, 21, 32, 33, 43, 59, 61, 64, 76, 77, 135, 140, 150, 152, 175, 177, 210, 247
- Schizotypal personality disorder: 1, 5, 6, 11, 21, 32, 33, 49, 53, 61, 64, 67, 76, 77, 140, 152, 177, 196

High scoring females

- Borderline personality disorder: 13, 41, 84, 85, 128, 138, 162, 168, 170, 174, 184, 206, 217, 219, 234, 241, 245, 246
- Eating disorders: 13, 41, 70, 84, 89, 96, 109, 121, 138, 156, 162, 170, 185, 205, 212, 219, 211, 214
- Schizotypal personality disorder: 13, 25, 41, 96, 102, 116, 118, 139, 170, 191, 206, 219, 224, 230, 240, 241, 245, 246

Low scoring females

- Borderline personality disorder: 2, 8, 36, 63, 66, 75, 86, 87, 111, 120, 134, 155, 166, 171, 172, 179, 189, 236,
- Eating disorders: 10, 56, 71, 75, 80, 86, 99, 115, 124, 144, 155, 171, 174, 190, 223, 226, 228, 253,
- Schizotypal personality disorder: 8, 12, 63, 70, 101, 109, 110, 111, 134, 155, 166, 172, 179, 185, 189, 190, 193, 214,

Sex differences between composite scores

There are a number of disorders in which the average scores of male and female individuals creating the composite images show a large discrepancy.

Trait	Sex	High M	High SD	Low M	Low SD
Depression (Photo shoot A)	Female	34.73	5.60	2.53	1.46
	Male	27.73	9.07	3.13	2.23
Depression (Photo shoot B)	Female	30.40	8.40	5.40	1.74
	Male	25.40	7.69	6.40	2.36
SPD (Photo shoot B)	Female	44.50	8.02	5.33	2.82
	Male	39.72	9.81	11.05	5.72
SPD (Photo shoot C)	Female	15.89	2.12	1.28	0.90
	Male	12.82	2.69	3.39	1.70
BPD	Female	45.72	19.05	1.89	1.28
	Male	19.89	2.39	7.26	2.12
ASD	Female	25.00	4.49	8.33	1.35
	Male	21.47	4.16	9.40	3.78

The modest sex differences in the average of high scores of depression may be explained by previously identified higher scores in female samples (Baron & Perron, 1986). In addition females have been found to score significantly higher than men on the SPQ (Raine, 1992).

The remarkable sex differences in scores are found in BPD and ASD. In BPD the high female mean score is almost double that of males, and the low score six points below the male mean. Although an increased prevalence in females has previously been documented (Lieb, Zanarini, Schmahl, Linehan & Bohnus, 2004), it is likely that this discrepancy is driven by the low sample size of males within the photo shoots, for example photo shoot C had 141 females and only 90 males. In addition this sample size further reduced by removing males with facial hair.

In ASD the average female score is larger than for males, this contradicts previous findings illustrating a marked dominance for males on this scale (Baron-Cohen et al., 2001). Again it is likely the reduced sample size is likely to be the cause of a low top mean, as males had larger scores overall (Male range 3-41, $M = 17.31$, $SD = 6.52$, Female Range 6-35, $M = 16.30$, $SD = 5.59$).

Correlations between questionnaires

This section reports the correlations between participant's scores on each questionnaire within each photo shoot. As expected from previous research identifying a common trend of low Extraversion and high Neuroticism scores in individuals with high scores in a range of mental health traits, there were positive correlations between mental health traits and Neuroticism and negative with Extraversion across all photo shoots. The one disorder which consistently defied this correlation was bipolar disorder, demonstrating a positive correlation with both Extraversion and Neuroticism in photo shoot B and C. Additional disorder which did not match the pattern were Eating disorders (Photo shoot C) and ADHD (Photo shoot B).

In addition there were consistent positive correlations between mental health traits. Perhaps this is unsurprisingly when each trait exhibits a very similar personality style. The correlations between traits has an impact on the production of composite images used throughout the thesis as individuals scoring in the extremes for one trait are likely to do so on others, reflected in the overlap between individuals across composite images demonstrated above.

The high correlations between scores across mental health traits and a consistent pattern of correlations with personality trait lends support for the argument that the identification of mental health traits may be a reflection of a single continuum of social desirability. That is, that perceptions of individual traits may not be distinct from each other but a reflection of a general perception of social desirability (or in terms of Oosterof and Todorov, 2008, dominance and trustworthiness) which is utilised when asked to distinguish between high and low images.

Photo shoot A

	Depression	**E**	**A**	**C**	**N**	**O**	**Dominance**	**ASD**	**SF12 - physical health**	**SF 12-MH**
Depression		-.134	-.048	-.206	.595	.051	.109	.413	.055	-.642
E	-.134		.176	-.026	-.234	.082	.181	-.500	-.063	.270
A	-.048	.176		.168	.137	.023	-.223	-.198	-.045	.033
C	-.206	-.026	.168		-.127	-.006	-.058	.021	-.039	.199
N	.595	-.234	.137	-.127		-.034	.0003	.398	-.011	-.596
O	.051	.082	.023	-.006	-.034		.211	-.1	-.073	-.089
Dominance	.109	.181	-.223	-.058	.0003	.211		.154	-.005	-.077
ASD	.413	-.500	-.198	.021	.398	-.100	.154		-.090	-.383
SF12 - physical health	.055	-.063	-.045	-.039	-.011	-.073	-.005	-.090		-.280
SF12 - MH	-.642	.270	.033	.199	-.596	-.089	-.077	-.383	-.280	

Photo shoot B

	Depression	A	C	E	N	O	Dom	SF12-Physical	SF12-MH	Pos PANAS	Neg PANAS	Schiz	ADHD	ASD	Anxiety	OCD	Bipolar
Depression		-.259	-.117	-.003	-.295	.160	.208	.044	.086	.045	.016	.028	.145	.022	-.098	.022	.070
A	-.259		.115	.167	-.005	.150	-.306	.013	.057	.102	-.034	-.148	.007	-.202	-.091	-.085	.071
C	-.117	.115		-.065	-.085	-.047	-.071	.117	.122	.082	-.128	-.096	-.378	.164	-.184	-.111	-.123
E	-.003	.167	-.065		-.197	.241	.183	.030	.248	.091	-.212	-.209	.115	-.298	-.273	-.088	.194
N	-.295	-.005	-.085	-.197		-.171	.031	-.146	-.455	-.186	.309	.439	.203	.337	.614	.358	.197
O	.160	.150	-.047	.241	-.171		.041	-.096	.105	.056	.015	.031	.054	-.106	-.084	.046	.077
Dom	.208	-.306	-.071	.183	.031	-.041		.062	.014	.107	.096	.193	.109	.144	-.004	.264	.196
SF12-Physical	.044	.013	.117	.030	-.146	-.096	.062		.168	.006	-.042	-.121	-.034	-.139	-.128	-.067	.014
SF12-MH	.086	.057	.122	.248	-.455	.105	.014	.168		.094	-.339	-.450	-.245	-.206	-.612	-.286	-.174
Pos PANAS	.045	.102	.082	.091	-.186	.056	.107	.006	.094		.150	-.050	-.136	.102	-.214	.012	.062
Neg PANAS	.016	-.034	-.128	-.212	.309	.015	.096	-.042	-.339	.150		.439	.150	.282	.556	.301	.139
Schizotypy	.028	-1.48	-.096	-.209	.439	.031	.193	-.121	-.450	-.050	.439		.298	.468	.628	.552	.357
ADHD	.145	.007	-.378	.115	.203	.054	.109	-.034	-.245	-.136	.150	.298		.063	.312	.263	.294
ASD	.022	-.202	.164	-.298	.337	-.106	.144	-.1139	-.206	.102	.282	.468	.063		.395	.432	.109
Anxiety	-.098	-.091	-.184	-.273	.615	-.084	-.004	-.128	-.612	-.214	.556	.628	.312	.395		.456	.211
OCD	.022	-.085	-.111	-.088	.358	.046	.264	-.067	-.286	.012	.301	.552	.263	.432	.456		.414
Bipolar	.070	.071	-.123	.194	.197	.077	.196	.014	-.174	.062	.139	.357	.294	.109	.211	.414	

Photo shoot C

	A	C	E	N	O	Flourishing	Wellbeing	ASD	SOI	Bipolar	Depression	ED	SF12 Physical	SF12 MH	SPD	BPD
A		.086	.251	.140	.204	.287	.383	.010	-.012	.141	.156	.035	.093	.075	.081	.074
C	-.086		.046	.013	-.034	.325	.135	-.058	-.079	-.112	.018	.060	.100	.119	-.073	-.080
E	.251	.046		-.056	.104	.301	.106	-.071	.139	.177	-.015	.106	.071	.119	-.082	-.094
N	.140	.013	.013		.068	-.132	.147	.301	.071	.104	.343	.088	.007	-.248	.209	.405
O	.204	-.034	-.034	.068		.146	-.021	.004	.129	.112	.066	.122	.153	.123	.152	.072
Flourishing	.287	.325	.325	-.132	.146		.337	-.062	.027	.001	-.167	-.037	.088	.240	-.092	-.251
Wellbeing	.383	.135	.135	.147	-.021	.337		.147	.034	.072	.108	.067	.183	.113	.031	.144
ASD	.010	-.058	-.058	.301	.004	-.062	.147		.064	.141	.217	.059	-.053	-.136	.187	.292
SOI	-.012	-.079	-.079	.071	.129	.027	.034	.064		.231	.148	.101	.132	.064	.075	.163
Bipolar	.141	-.112	.177	.104	.112	.001	.072	.141	.231		.267	.138	-.054	-.034	.367	.158
Depression	.156	.018	-.015	.343	.066	-.167	.108	.217	.148	.267		.321	.048	-.186	.424	.658
ED	.035	.060	.106	.088	.122	-.037	.067	.059	.101	.138	.321		.033	.023	.211	.386
SF12 Physical	.093	.100	.071	.007	.153	.088	.183	-.053	.132	-.054	.048	.033		.654	.121	.122
SF12 MH	.075	.119	.119	-.248	.123	.240	.113	-.136	.064	-.034	-.186	.023	.654		-.053	-.276
SPD	.081	-.073	-.082	.209	.152	-.092	.031	.187	.075	.367	.424	.211	.121	-.053		.424
BPD	.074	-.080	-.094	.405	.072	-.251	.144	.292	.163	.158	.658	.386	.122	-.276	.424	