A *Different 'Different':* *The Female Presentation of Autism Spectrum Disorder and Implications for Detection and Diagnosis*

by

Joanna M. Tsirgiotis
BPsych(Hons)

Thesis
Submitted to Flinders University
in partial fulfilment of the requirements
for the degree of

Doctor of Philosophy (Clinical Psychology)

College of Education, Psychology, and Social Work

November 2020

Table of Contents

Abstract

Autism spectrum disorder (ASD) is diagnosed more commonly among males than females (Loomes et al., 2017). Although females may be less likely to develop the condition due to biogenetic protective factors (Robinson et al., 2013), growing evidence suggests that a large number of autistic females remain undiagnosed and are thus unable to access the specialised support they may require (Hull et al., 2020). This thesis examines potential reasons for the underdiagnosis of females with ASD, with a particular focus on fine-grained behavioural differences between males and females and possible bias related to the interpretation of autistic behaviours.

In order to investigate the specific behaviours and domains in which males and females differ in the severity of their ASD difficulties, in Study 1, I analysed item-level profiles of 777 children using the Childhood Autism Rating Scale, 2nd Edition (CARS2; Scholper et al., 2010) or Gilliam Autism Rating Scale, 3rd Edition (GARS-3; Gilliam, 2014). Males demonstrated greater difficulty in six CARS2-ST items and seven specific behaviours on the GARS-3, most of which reflected specific restricted and repetitive behaviours. Across all instruments, the only area in which females showed greater difficulty was fear or nervousness (CARS2-ST). No meaningful differences emerged from the CARS2-HF analysis. On the items where males showed greater difficulty, females were more likely to present with developmentally typical behaviour.

Study 2 was comprised of two parts, each of which addressed the issue of why some females with many ASD traits are not diagnosed with ASD when they present for assessment. In Study 2a, I explored changes in the presentations of 12 girls who were diagnosed with ASD only after an initial negative result. A number of specific social difficulties emerged between assessments, particularly in the content of conversation. Further, there was a meaningfully higher probability that they would meet Criterion B2, insistence on sameness,

routines, and ritualised behaviour at the time of the second assessment. In Study 2b, the presentations of both males and females who were either diagnosed with ASD (n = 156) or not diagnosed with ASD despite many ASD traits, being suspected of having ASD, and being referred for assessment (n = 78), were compared. Two important contributions of Study 2b were: (a) the inclusion of females whose presentation deviates from the classic male conceptualisation, did not meet criteria and remained undiagnosed (often excluded from the research to date), and (b) consideration of diagnostic data from different report sources: parent report, diagnostic observations, and teacher report. Results showed that females were less likely than males to meet Criterion B3 (restricted interests), and this was especially the case for subclinical (non-ASD) females. Indeed, of all criteria, females who presented due to ASD concerns and were either diagnosed or not diagnosed, were least likely to meet Criterion B3. Evidence of sex/gender specific restricted interests and stereotypical behaviours was found. Further, teachers and diagnosticians were less likely to report concern for females than for males. Importantly, many behaviours differed in the extent to which they predicted the ASD diagnostic result for males and females, perhaps suggesting that sex/gender influences how ASD-related behaviours are perceived.

In Study 3, 47 ASD diagnosticians were presented with two hypothetical case studies (one male ASD presentation and one female ASD presentation), and the sex/gender of the child described was randomly assigned within each. Diagnosticians reported greater ASD symptom severity when female sex/gender pseudonyms were allocated to the case studies, but their confidence in ASD diagnosis was similar regardless of the sex/gender condition. Diagnosticians identified a large number of challenges associated with assessing females for ASD. Many of these related to sex/gender differences in ASD presentation and difficulties in detecting the presentation of females. Broadly, results provided new insight into why ASD may be under-identified and underdiagnosed among females and provide evidence to support

a broader and/or clearer and more flexible conceptualisation of ASD in order to better reflect the difficulties of autistic females and promote greater diagnostic certainty.

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Joanna M. Tsirgiotis

21 August 2020

Acknowledgements

To my supervisors, Professor Robyn Young and Associate Professor Nathan Weber, it has been an honour and privilege to learn from you throughout my PhD. I am so grateful for the unwavering support you have provided me and the skills and knowledge I have gained. Thank you for your mentorship, patience, wisdom, and above all, your belief in me.

My parents, Kathy and Evan, sister, Philippa and closest friends have been pillars of undying encouragement and strength. Thank you for sharing the tears and joys with me every step of the way. To my Aunty Eleni, Nouna Betty, and extended family, thank you all for your encouragement and good humour. To my clinical psychology cohort, where would we be without each other's support? I am so grateful to have shared this journey with you.

Thank you to the Headstart clinicians who have mentored and supported me to develop my understanding and skills in ASD assessment. My thesis would be far poorer were it not for my experience with you. To Bethany Phipps and Jeremy Lim, thank you for your assistance with statistics and interrater reliability checks. Special thanks also go to Professor Anna Chur-Hansen for her advice on qualitative data analysis, and Dr Alliyza Lim for her guidance and support. Finally, thank you to the organisations who assisted me in recruiting for my studies and to the diagnosticians and families who generously offered their time and insights in pursuit of a deeper understanding of ASD.

I dedicate this thesis to my grandparents. It was your selflessness, dedication and love which brought me an abundance of opportunities that you did not have. Despite limited schooling opportunities, you have forged and supported your children's successes by recognising the value of education. We are forever indebted to you all.

I acknowledge the support of an Australian Government Research Training Program Scholarship throughout my candidature. A professional editor has *not* been deployed in the production of this thesis.

Chapter 1: Introduction

Overview

The male preponderance in autism spectrum disorder (ASD) has been apparent since the two seminal case studies describing autistic characteristics in children (Asperger, 1944; Kanner, 1943). Leo Kanner, the first to describe such characteristics, presented 11 case studies of which only three described female children. All of the four cases presented in Hans Asperger's seminal work of *Autistic Psychopathy* [sic] were male children. Asperger commented that, "It is fascinating to note that the autistic children we have seen are almost exclusively boys," adding that some girls had "contact disturbances which were reminiscent of autism,'' but that there were none with "fully formed" or "fully fledged" characteristics (Asperger, 1944; Frith, 1991, pp. 84-85). Asperger's writing highlights that, even at the time that ASD was first being described and conceptualised, the apparent imbalance in prevalence between males and females was recognised as an important, yet unexplained feature.

To date, this male preponderance remains among the most consistent features of ASD, with diagnosed males considerably outnumbering females at all developmental stages. A recent meta-analysis of epidemiological research concluded that there are between four and five times as many males diagnosed with ASD as females (Loomes et al., 2017).[1] However, the size of this difference is not consistent across the spectrum of functioning and intellectual ability. In the absence of intellectual disability, the ratio of males to females diagnosed with ASD has been found to be as high as 10:1 (Fombonne, 2009) in contrast to 2:1 amongst individuals with co-occurring intellectual disability (Fombonne, 2005).

[1] Identity first (autistic girl/boy/[wo]man) and person first (girl/boy/[wo]man with ASD/autism) language will be used interchangeably to reflect the different preferences of members of the ASD community (Kenny et al., 2016).

Although there is evidence to suggest that females are less likely to develop ASD than males due to biogenetic and aetiological factors (Ferri et al., 2018), it is also likely that the size of the sex/gender[2, 3] discrepancy is exaggerated due to under-identification and underdiagnosis of the condition in females (Kirkovski et al., 2013; Rutter et al., 2003). As a result, the true prevalence ratio (inclusive of autistic females without a formal diagnosis) is unknown. Evidence supporting the existence of a distinctive female ASD presentation (otherwise referred to as the *female ASD phenotype*) is growing (Hull & Mandy, 2017).[4] As it may bear subtle quantitative and qualitative differences to the typical or 'classic'[5] ASD presentation, the female phenotype may be incongruent with current assessment instruments, diagnostic criteria, and clinical understanding, meaning that those who most embody this phenotype may not be diagnosed and may thus be excluded from the majority of research to date (Lai, Lombardo, et al., 2015). This thesis will examine fine-grained sex/gender differences in the presentations of autistic children (aged between 2 years and 17 years 11

---

[2] *Sex* refers to biological characteristics differentiating males and females and *gender* to socially constructed roles and attributes viewed as normative for a particular sex (American Psychiatric Association, 2011). Gendered socialisation begins at birth and resultantly, biological sex and socialised gender are not easily separated, mutually informing an individual's identity. Therefore, as proposed by Springer et al. (2011) and recommended by Lai et al. (2015), the term *sex/gender* will be used in this thesis to reflect the overlap between both constructs (unless otherwise stated).

[3] Binary notions of sex and gender exclude the lived experiences of a number of individuals, including those who are intersex or identify as transgender or gender diverse. However, issues specifically pertaining to sex and gender diverse individuals, while recognised, fall beyond the scope of this thesis.

[4] It is acknowledged that not all autistic females will present with a 'female' ASD phenotype and equally, that not all males will present with a 'male' ASD phenotype. This thesis is concerned with issues pertaining to underdiagnosis of autistic females, which may be magnified as their presentation deviates from the 'classic' ASD presentation. The issue of males presenting with female ASD features will not be examined directly.

[5] The term *classic ASD* is used to refer to the typical perception of the ASD presentation.

months), including females who fail to meet ASD diagnostic criteria at their first assessment and those with many autistic traits but who do not fully meet criteria. More broadly, I will explore issues pertaining to the possible under-identification and underdiagnosis of autism in girls and women, with the general purpose of aiding clinicians' understanding of how ASD manifests in females and thus facilitating its more timely and accurate identification.

Reasons for the Imbalance in Diagnostic Prevalence

Two broad hypotheses have been put forward to account for the apparent male preponderance in ASD. The first is that biological and genetic factors may 'protect' females from developing ASD as readily as males, and therefore the imbalanced ratios reflect a reality that, across the entire spectrum, approximately four to five times as many males have ASD compared to females. The second hypothesis is that ASD is underdiagnosed in females. Given the evidence in support of each hypothesis, it is likely that they are not mutually exclusive (Chen et al., 2020). That is, it is possible that although females may be less likely to develop ASD, there may be a subset of autistic females who remain unidentified and undiagnosed.

Biological and Genetic Factors

It has been argued that females may be less likely than males to develop ASD as a result of biogenetic factors (Ferri et al., 2018). Although multiple theories have been outlined to account for this possibility, among the two most persuasive arguments are the Extreme Male Brain theory (Baron-Cohen, 2002; Baron-Cohen & Hammer, 1997) and the Female Protective Effect (Jacquemont et al., 2014).

The Extreme Male Brain (EMB) theory purports that females are less likely to develop ASD due to biological differences between typically developing males and females. Specifically, the theory suggests that in the typically developing population, males have better-developed systemising skills (i.e., identifying patterns, rules, and details), females have

superior empathising ability (i.e., identifying, understanding and appropriately responding to emotions in others), and that autistic people tend to process the world in a highly analytical and systematic (male) manner. While there is some support for this theory (Teatero & Netley, 2013), it cannot fully account for the nature of differences in ASD between males and females (Bejerot et al., 2012; Blakemore et al., 2004), particularly given methodological limitations in the initial research and the simplistic nature of the theory (Ridley, 2019). Specifically, the EMB theory cannot explain whether these differences reflect gendered socialisation or biological factors, such as differences in genetic aetiology (Constantino & Todd, 2003).

The genetic aetiology of ASD is currently understood to be dependent on many mutations (including deletions and duplications), rather than one single gene or interaction, which may contribute to the heterogeneity in the disorder's expression between individuals (Sandin et al., 2014). The Female Protective Effect theory suggests that a generally higher level of genetic liability is necessary for females to develop neurodevelopmental disorders (Robinson et al., 2013), and thus ASD characteristics[6] are less pronounced in females who have genetic liability equal to males. In support of this theory, a study by Jacquemont et al. (2014) found that females with ASD have more ASD-related genetic mutations than males with ASD (see Werling & Geschwind, 2013). Together, the Extreme Male Brain and Female Protective Effect theories suggest that a higher genetic threshold is necessary for females to develop ASD.

[6] In this thesis, every effort is made to avoid pathologising or disempowering language in referring to the experiences of autistic people. For example, where the broader literature may refer to autistic *impairment*, the terms *atypicality* or *ASD-related difficulty* are used here. Similarly, *behavioural impairments* are referred to as ASD *characteristics*, *behaviours*, *features,* or *traits*.

Underdiagnosis of ASD in Females

While females may be less likely to develop ASD than males, it is also possible that females who do develop the condition are less likely to be identified. In contrast to the aforementioned ratio of individuals *diagnosed* with ASD (four to five males for every female), active identification or screening of cases of ASD in the general population, regardless of existing diagnosis, has estimated that the true prevalence ratio is 3.25 males for every female (Loomes et al., 2017). Further, a predictive model recently constructed from population data estimated that 39% more girls might be expected to be diagnosed with ASD (Barnard-Brak et al., 2019).

Evidence suggests that cognitively able females with ASD are, on average, diagnosed later than their male counterparts, if their ASD is indeed identified. For example, Siklos and Kerns (2007) found that in a sample of 56 autistic children and adolescents, girls were diagnosed with ASD on average 18 months later than boys, despite having visited the same number of professionals. This is consistent with a larger study of 2,000 individuals with ASD, which found that girls were diagnosed later than boys despite no significant difference in the age at which parental concern was first raised (Begeer et al., 2013). The delay in the time taken to diagnose ASD in girls suggests that the diagnostic process may be more challenging. We can only speculate as to why this delay may occur, but it may be due to (a) sex/gender differences in how the disorder presents, and (b) the familiarity of professionals with the female ASD presentation. These possibilities are not mutually exclusive, and both will constitute key foci of this thesis.

While some females will be diagnosed with ASD, albeit for many with some delay, others may continue to be considered subclinical and/or alternative diagnoses may be considered. By way of illustration, Wilson et al. (2016) found that in their sample of 1,244 adults referred for ASD assessment, males were more likely to receive an ASD diagnosis, but

females were more likely to receive *partial* ASD diagnoses (such as pervasive developmental disorder not otherwise specified, or social communication disorder).[7] Similarly, Ratto et al. (2018) found that in an age and Intelligence Quotient (IQ) matched sample of 228 children diagnosed with ASD (114 girls), girls with higher cognitive ability were less likely to meet diagnostic criteria for ASD despite experiencing more severe autistic traits and difficulty in parent-reported adaptive skills. Other evidence suggests that in order for females to receive an ASD diagnosis, they must display greater intellectual and/or behavioural difficulties (Dworzynski et al., 2012), emotional challenges (Duvekot et al., 2016), and more severe ASD characteristics (Russell et al., 2011) than males. Indeed, parent-reported repetitive and restricted behaviour difficulties were more likely to lead to an ASD diagnosis in males than females (Duvekot et al., 2016). As such, the sex/gender of the individual may predict whether an ASD diagnosis is made. Multiple variables may contribute to whether the ASD assessments result in a diagnosis (e.g., intelligence, age, and the severity of behavioural difficulties, emotional challenges, and ASD characteristics), but these may interact with sex/gender (i.e., these influences may differ in strength according to the sex/gender of the child).

Sex/Gender Differences in ASD Presentations

ASD is currently conceptualised and diagnosed based upon behavioural features within two broad criteria: impairment and delay in the development of social communication and social interaction capacities, and the presence of restricted and repetitive patterns of behaviour, interests or activities (American Psychiatric Association, 2013).[8] The current conceptualisation of ASD as a neurodevelopmental spectrum highlights the variability in the

[7] In social communication disorder, only the social communication criteria of ASD are met (i.e., there is an absence of repetitive and restricted behaviour).

[8] The diagnostic criteria for ASD according to the *Diagnostic and Statistical Manual of Mental Disorders 5th Edition* (DSM-5; American Psychiatric Association, 2013) are presented in Appendix A.

severity and manifestation of intellectual and functional difficulties between affected individuals, and the vast heterogeneity in cognitive and behavioural presentations. As a result of this heterogeneity, no overarching explanation or theory of ASD has been accepted as adequately accounting for all of its features and symptom expressions (Happé et al., 2006). Heterogeneity in the expression of ASD between males and females has received special attention only within the last decade.

Perhaps as a result of the male preponderance in ASD, the majority of investigations into the features of ASD have traditionally been conducted with androcentric and clinical participant samples which are consistent with the imbalanced prevalence ratio, or exclude females entirely (Kreiser & White, 2014; Rutter et al., 2003). As a result, the very nosology and conceptualisation of ASD has been based upon this androcentric literature. Thus, if a distinct and sex/gender-specific, female presentation of ASD exists, it is possible that the current diagnostic criteria and assessment instruments do not sufficiently capture it (Lai, Lombardo, et al., 2015; van Wijngaarden-Cremers et al., 2014). Therefore, females who embody this phenotype may remain undiagnosed and 'fly under the diagnostic radar.' An important contribution of this thesis is the consideration of presentations of males and females who, despite demonstrating many ASD behaviours, do not fully meet the diagnostic criteria for ASD, and were previously not represented in the literature.

Existing research surrounding the possibility of a female ASD phenotype has generally followed two lines of enquiry. The first of these examines quantitative differences (i.e., severity) in the core ASD features between clinically derived samples of males and females diagnosed with ASD (i.e., difficulties with social communication and social interaction; and restricted and repetitive behaviours and interests; RRBIs) as measured by standardised assessment instruments. The second of these is more limited and explores

qualitative differences in the manifestation or expression of ASD features between males and females. Both of these lines of enquiry will now be addressed in turn.

Quantitative Differences in ASD Features Between Males and Females

Research findings pertaining to differences in social and communicative difficulties in males and females have been largely inconsistent. While some studies suggest that females with ASD show significantly less severe (Lai et al., 2011; McLennan et al., 1993), others have shown them to be more severe (Hartley & Sikora, 2009), and others suggest they have equally severe social and communicative difficulties compared to autistic males (Andersson et al., 2013; Mandy et al., 2011). It is therefore not surprising that van Wijngaarden-Cremers and colleagues' meta-analysis of 22 studies examining differences between males and females in the core features of ASD (2014) found no statistically significant discrepancies in the overall severity of difficulties in social behaviour or communication in any age group.

Nevertheless, studies that have compared RRBIs in clinically derived samples of males and females with ASD have more consistently found fewer and less pronounced RRBIs among females in all age groups beyond toddlerhood (e.g., Tillmann et al., 2018; van Wijngaarden-Cremers et al., 2014). Van Wijngaarden-Cremers and colleagues (2014) found that the difference between RRBI severity in males and females with ASD was statistically significant overall (std. mean diff. = 0.51, $CI_{95\%}$ = [0.21, 0.81], $p < .001$). The evidence of a sex/gender difference was strongest in adolescents (std. mean diff. = 0.69, $CI_{95\%}$ = [0.33, 1.05], $p < .001$), followed by children (std. mean diff. = 0.19, $CI_{95\%}$ = [0.06, 0.32], $p < .001$), and adults (std. mean diff. = 0.47, $CI_{95\%}$ = [0.03, 0.92], $p < .05$), but not for toddlers. Consistent with this, greater RRBI difficulties in males have been found in individual studies of adults (Lai et al., 2012) and children (Bölte et al., 2011; Mandy et al., 2011). Studies of toddlers with ASD have produced inconsistent results, with some showing greater atypicality

amongst males (Hartley & Sikora, 2009) and others showing no significant sex/gender differences (Carter et al., 2007).

Multiple methodological issues hamper our ability to make meaningful conclusions across studies examining relative symptom severity. Specifically, differences in ascertainment procedures (i.e., whether clinical or population samples were used) and the age and intellectual level of participants affect our ability to draw conclusions about sex/gender differences (Kirkovski et al., 2013; Rivet & Matson, 2011b; van Wijngaarden-Cremers et al., 2014). Additionally, participant samples, particularly of females, have typically remained small.

Qualitative Differences in ASD Features Between Males and Females

Autobiographical works (Holliday-Willey, 2015), accounts of clinicians (Attwood, 2007), quantitative studies, and qualitative investigations based on interviews with autistic female clients and their families have also suggested that ASD may manifest qualitatively differently in females compared to males. That is, the *expression*, rather than severity of ASD features may differ by sex/gender.

Although there is a great deal of symptom overlap, some features of social interactions among females with ASD have been identified as being qualitatively different from those of the male perception of ASD. Motivation for social interaction has been shown to be higher in adolescent girls with ASD than adolescent boys with ASD (Head et al., 2014; Sedgewick et al., 2016) and attention to social stimuli (e.g., faces), commonly used to operationalise social motivation in ASD, has been shown to be higher in autistic primary-school age girls compared to their male counterparts (Harrop et al., 2018). Autistic females have described themselves as eager for friendships, desiring social contact with peers, and being aware of a need for social interaction (Tierney et al., 2016; Vine Foggo & Webster, 2017). Research examining the specific ways in which girls and boys come to meet the

criteria for ASD has reported distinctive patterns in social domains (e.g., girls were found to be significantly less impaired than boys in social and emotional reciprocity, and in their ability to share interests and to initiate but not maintain friendships; Hiller et al., 2014). Autistic females may also experience more difficulty managing conflict in social relationships than both autistic males and typically developing females during adolescence (Sedgewick et al., 2019). Other social features, such as a tendency to be perceived as shy, be controlling in play with peers, and maintaining only one (or a small number) of close or *intensive* friendships have been described by adult women with ASD (Holliday-Willey, 2015) and clinicians (Attwood et al., 2006; Lai, Lombardo, et al., 2015).

As outlined above, there is evidence to suggest that females with ASD have fewer and less severe RRBIs than males with ASD. However, rather than being less severe, it is also possible that females' RRBI behaviours may be less clinically recognisable because they are qualitatively distinct from those of males (Kopp & Gillberg, 1992). Restricted interests are among the RRBIs most studied thus far, with evidence suggesting that these may be less obviously atypical or more developmentally or gender appropriate among females (Attwood, 2007; Hiller et al., 2014; Kopp & Gillberg, 1992). In contrast, Hiller et al. (2014) found that girls were more likely to have seemingly random restricted interests or interests in toys (rather than screens or wheeled vehicles). Indeed, it has been suggested that females may develop restricted interests in social or relational subjects, such as with one particular friend, which may superficially conceal socio-communicative difficulties and not be considered a restricted interest (Attwood et al., 2006; Kopp & Gillberg, 1992). In addition to more socially oriented interests, females may express their interests differently to males, particularly in more social ways (e.g., volunteering with animals rather than collecting animal figurines; McFayden et al., 2018). It is also possible that, although interests may be atypical in intensity for males and females, the less atypical orientation of female restricted interests may result in

less day-to-day disruption for the autistic individual and her family (Hull et al., 2020). Thus, restricted interests may be less likely to be reported as unusual by parents, or their functional impact may be underestimated.

Equally, stereotypical behaviour profiles may be influenced by sex/gender. A recent study examined sex/gender differences in specific RRBI behaviours through an item-level analysis of children's scores on the Repetitive Behavior Scale-Revised (RBS-R; Lam & Aman, 2007), and found more severe stereotypical behaviour amongst boys (e.g., hand and finger mannerisms and object use), but elevated hoarding, self-injurious behaviour, and insistence on sameness amongst girls (Antezana et al., 2018). However, differences have not been consistent within the sensory hyper-/hypo-sensitivity domain, with Bitsika et al. (2018) finding no significant sex/gender differences in a fine-grained analysis of children's scores on the Sensory Profile (Dunn, 1999). As such, sex/gender differences may exist in some, but not all, ASD behavioural domains.

The possibility that males and females with ASD differ in how certain behaviours are expressed has important implications for the identification of the disorder in females. Specifically, qualitative differences between males and females may contribute to females' under-identification due to possible bias toward the typical 'male' behaviour manifestations in commonly used assessment instruments, the expectations of clinicians, and how overtly atypical behaviours appear in any given environment. Each of these possibilities will be discussed in greater depth below. To date, the literature has generally failed to explore symptom manifestation in females with narrowly sub-threshold ASD, or those who are diagnosed following an initial negative result. Therefore, it is unclear which criteria are commonly unmet or which behaviours are less likely to be demonstrated (or most difficult to discern as atypical), and consequently which criteria are least sensitive to females' difficulties.

There has been suggestion that across both sexes, those who meet the criteria for ASD in the *Diagnostic and Statistical Manual of Mental Disorders, Text Revised Fourth Edition* (DSM-IV-TR; American Psychiatric Association, 2000) but fail to meet the criteria for ASD outlined in the fifth edition (DSM-5; American Psychiatric Association, 2013) do so because they do not fully meet requirements in the socio-communication domain (Young & Rodi, 2014). These authors argue that the DSM-5 criteria may be overly exclusive in these domains. Although yet to be empirically demonstrated, it is possible that these criteria are also those least likely to be met by females. However, given the evidence for both less severe RRBI atypicality and less overtly atypical restricted interests in females diagnosed, it is also possible that females will fail to meet these criteria due to difficulties not reaching clinical significance. An important contribution of this thesis will be to determine which ASD behaviours are less likely to be present in females, and/or how behaviours may manifest differently within each of the criteria.

Why Might Some Females Fall Under the ASD Diagnostic Radar?

In addition to differences in the severity and qualitative manifestation of ASD difficulties between males and females, autistic females may remain under-identified due to camouflaging ASD related difficulties. Moreover, different developmental trajectories may influence when difficulties emerge. Differences in the nature and prevalence of co-occurring psychiatric conditions and clinician biases and/or challenges associated with existing assessment procedures may also result in under-identification. Each of these possibilities and their implications will now be discussed in turn.

Social Camouflaging

One of the theories with the most traction in explaining why ASD may be under-identified in females, or why identification and diagnosis may be delayed, is known as *social camouflaging* (Wing, 1981). Social camouflaging was first described by Wing in 1981 but

has only recently been clearly defined and operationalised. As social camouflaging may contribute to both quantitative and qualitative differences in ASD presentation between males and females and will be important in interpreting the results presented in this thesis, it will be discussed in some depth here.

Social camouflaging is presently understood as a combination of learned socio-communicative behaviours used to disguise and compensate for ASD-related difficulties (Hull et al., 2017; Lai et al., 2016), and as such, may constitute an important social coping strategy for individuals with ASD (Attwood, 2007; Kopp & Gillberg, 1992; Wing, 1981). In their recent investigations into the camouflaging experiences of adults with ASD and construction of a self-report camouflaging questionnaire (Camouflaging Autistic Traits Questionnaire; CAT-Q), Hull et al. (2017; 2018) identified three distinct groups of behaviours that comprise camouflaging; masking, compensation, and assimilation. *Masking* behaviours aim to conceal or suppress ASD characteristics, and *compensation* behaviours include strategies such as imitation and mimicry (e.g., topics of discussion, tone of voice, choice of clothing) that are used to bridge socio-communicative gaps with typically developing peers. Finally, *assimilation* describes attempts to blend into social situations which cause the individual discomfort, without this discomfort becoming apparent to others. Although the development of these behaviours has not been well explored, it is possible that they may be partly learned from peers, by watching television or from literature. However, who develops these and what skills are required to develop camouflaging remains unknown.

While male and female adults with ASD have reported engaging in camouflaging (Hull et al., 2020), it appears to be particularly common among cognitively able autistic females and is therefore considered an important feature of the female ASD presentation (Cassidy et al., 2018; Hull et al., 2020; Schuck et al., 2019; Wood-Downie et al., 2020). Recent evidence has recorded significantly higher scores amongst females compared to males

on both discrepancy-based measures of camouflaging (i.e., between 'external' presentation and 'internal' or dispositional ASD traits; Lai et al., 2016), and direct measures of camouflaging behaviours (specifically in masking and assimilation behaviours, but not compensatory behaviours; Hull et al., 2019). Additionally, research into the first impressions made by autistic adults and children on naïve observers has found that although autistic females may be viewed significantly less positively (i.e., less socially competent) than non-autistic females, and approximately as positively as typically developing males, they are viewed significantly more positively than autistic males, despite equally severe ASD characteristics (Cage & Burton, 2019). Indeed, while Cola et al. (2020) found a significant association between clinician rated social difficulty and naïve observers' first impressions for autistic school age males, there was no significant association for females.

However, findings regarding sex/gender differences in camouflaging behaviours have been inconsistent. For example, while many studies have found higher camouflaging scores among autistic women compared to autistic men according to the CAT-Q (e.g., Hull et al., 2019) and other self-report measures (e.g., Cassidy et al., 2018), some have found no sex/gender differences in the likelihood of self-reported camouflaging among adults with ASD (Cage et al., 2018; Hull et al., 2017).

Although the reason(s) that many studies have found greater camouflaging among autistic females remains unknown, it is possible that greater social motivation (Sedgewick et al., 2016), higher cognitive ability, and/or fewer difficulties with social cognition and executive function (e.g., recognising emotions, imitation ability; Kothari et al., 2013; Lehnhardt et al., 2016; Livingston et al., 2019) may be contributing factors. A small number of qualitative investigations into the experiences of primary school age girls (Cook et al., 2017), female adolescents (Tierney et al., 2016), adult women (Baldwin & Costley, 2016; Bargiela et al., 2016), and adults of all sexes/genders (Cage & Troxell-Witman, 2019; Hull et

al., 2017), has identified that individuals' motivations for camouflaging primarily reflect the desire to have friends. Other motivating factors include reducing anxiety and embarrassment in social situations, shame regarding one's perceived inadequacies, and avoiding being bullied and 'standing out' as different to peers.

Some of the adolescent girls with ASD who were interviewed by Tierney et al. (2016) reported that their camouflaging was so successful that teachers were commonly surprised to learn of their ASD diagnosis. Similarly, the late-diagnosed adult participants recruited by Bargiela et al. (2016) illustrated that, at times, they camouflaged so well that they personally doubted that they had ASD. It is interesting to note that in qualitative interview studies, women with ASD (Bargiela et al., 2016) and the parents of girls with ASD (Navot et al., 2017; Rabbitte et al., 2017) reported their belief that females with ASD were different to males with ASD because of this ability to integrate with peers and "appear normal" (Holliday-Willey, 2015). It should be noted that much of the existing research pertaining to one's ability and motivation to engage in camouflaging is qualitative, and therefore, the generalisability of these findings remains unknown. Further, the development of camouflaging behaviours (i.e., when and in what ways they may emerge), has not been directly examined to date.

Blending-in, or having one's social difficulties not discerned by others, may be problematic in identifying ASD, both prior to specialist referral and during formal developmental assessment. Superficial social skills may mean that social difficulties are less likely to be detected, and therefore referral may be delayed or deemed unnecessary (Gould & Ashton-Smith, 2011; Kirkovski et al., 2013). Similarly, the presence of some social skills may be unduly construed as evidence that an ASD diagnosis is inappropriate (Lai & Baron-Cohen, 2015). It is therefore critical that questions are posed carefully during assessment. For example, rather than inquiring as to whether an individual shows empathy toward peers (a

behaviour that may be learned through observation and reinforced as being socially appropriate), it is important that the individual's desire, comfort, and flexibility in showing such empathy are assessed.

Some may argue that individuals who camouflage their ASD characteristics such that their ASD remains undetected may not require a diagnosis or associated support. However, qualitative investigations, autobiographical texts, and clinician accounts highlight that camouflaging comes at a cost, with many individuals experiencing exhaustion and identity confusion as a result of camouflaging, as well as anxiety around the successfulness of their efforts (Bargiela et al., 2016; Hull et al., 2017; Tierney et al., 2016). Indeed, Cassidy et al. (2018) identified camouflaging as a risk factor for suicidality amongst autistic adults, and camouflaging has been associated with greater mental health distress in women (Beck et al., 2020). Therefore, for many people who engage in camouflaging, appropriate (and ASD-informed) therapeutic support may be essential (Hull et al., 2017).

The Role of the Social and Physical Environment

Given that camouflaging may have adverse emotional consequences and is exhausting, it is thought to occur only in social environments. An important concept which has been applied by authors examining the environments within which camouflaging is likely to occur is *person-environment fit* (Hull et al., 2020; Lai & Baron-Cohen, 2015). For an autistic individual, the perceived incongruence between their genuine selves and the social environment may be a particularly motivating factor for camouflaging. This may be especially the case for individuals who are more sensitive to the incongruence, are more socially motivated or feel greater pressure to fit in with peers. The possibility that camouflaging, under-identification, and underdiagnosis of females with ASD are related to social and environmental contexts therefore warrants consideration. With regard to the nature of the female social environment, Dean, et al. (2016) argue that the *fluidity* of social

behaviour (i.e., the unstructured nature of activities, such as chatting or imaginative play) among typically developing females that enables girls with ASD to blend in and effectively conceal their social difficulties. In contrast, they observed that boys with ASD could be more easily identified (due to the structure of their play and isolation from the structured games played by their typically developing peers). In addition, it has been proposed that peers, parents, and/or teachers of girls are more likely to provide additional coaching or assistance in forming and maintaining relationships for females, particularly through scaffolding and providing opportunities to practice (Attwood et al., 2006; Tierney et al., 2016). As a result, girls may give the impression that they are managing socially.

Yet another factor that may increase the complexity of identifying and diagnosing ASD in girls is the possibility that their ASD difficulties may manifest differently, and to different extents, in different environments. In particular, their difficulties may be less discernible in the school environment (Attwood et al., 2006). Studies examining parent and teacher reports of the severity of ASD features and parent reports published in qualitative investigations have provided support for this hypothesis. It has been found that teachers generally report fewer ASD behaviours in children than their parents, and that this discrepancy may be greater in girls than in boys (Posserud et al., 2006). Despite no significant difference in the degree of social concern of parents of girls with ASD (Andersson et al., 2013) or greater concern for their daughters (Mandy et al., 2011) than parents of boys, teachers have been found to rate girls' social skills as better developed than their male counterparts (Hiller et al., 2014). Consistent with this, Hiller et al. (2014) found that in their sample, no concern regarding social skills was raised by teachers for 37% of girls with ASD compared with 5% of boys, despite them demonstrating enough global atypicality to meet the ASD diagnostic criteria.

Interviews with women diagnosed with ASD in adulthood (Bargiela et al., 2016) and parents of autistic girls (Rabbitte et al., 2017) have revealed how girls with ASD may be more adept at concealing their difficulties (e.g., emotional or social difficulties) at school than boys. Participants in Bargiela and colleagues' study (2016) described themselves as 'shy', 'passive', or 'good' and able to 'keep it together' at school, and then having regular 'meltdowns' at home. It is likely that this behaviour is a reflection of context specific camouflage, which becomes exhausting, and after which, it takes time to 'reset.' This may involve the expression of emotional distress in safe or private environments (Hull et al., 2017). As teachers are often the first to raise concern regarding a child's social development (which may lead to referral for ASD assessment), this suggests that girls' outward presentation at school may delay or prevent an ASD diagnosis from being sought and thus impede adequate support from being provided to the child and her family.

The Stage at which Social Difficulties Become Apparent

Yet another hypothesis suggests that girls' ASD characteristics may emerge at later developmental stages compared to males. It remains unclear as to whether (a) the onset of girls' social difficulties genuinely occurs later, or (b) existing social difficulties become more salient at later developmental stages (Kaat et al., 2020; Mandy et al., 2018). Consistent with the latter hypothesis, it has been suggested that camouflaging may be sufficient to mask and compensate for ASD related social difficulties during primary school for some girls with ASD, but this may become insufficient when the complexity of social interactions increases with adolescence (Hsiao et al., 2013; Tierney et al., 2016). For example, an autistic girl, able to maintain superficial friendships during primary school, may be less able to do so when the complexity of relationships increases with the onset of puberty and adolescence. Rather than the severity of the social difficulty increasing, the changes in the social environment may expose or exemplify existing social difficulties. Trajectories of autistic social traits in the

typically developing population suggest a more rapid escalation for females than males in early adolescence (age 10-16 years; Mandy et al., 2018). These possibilities may explain the apparent decrease in the ratio of males to females referred for assessment from 5:1 in childhood to 2:1 in adolescence and adulthood (Rutherford et al., 2016).

Related to the hypothesis that ASD difficulties may become more apparent at later developmental stages is the possibility that camouflaging behaviours have a detrimental effect on one's emotional wellbeing over time (Hull et al., 2017) and may lead to the development of co-occurring internalising difficulties such as depression and anxiety. Thus, females may come to clinical attention as a result of the secondary psychiatric conditions that may arise from living with undiagnosed ASD and without appropriate therapeutic support (see also Bargiela et al., 2016).

The presence of co-occurring psychiatric conditions may play a role in the underdiagnosis of females with ASD if (a) ASD traits are inappropriately ascribed to the pre-existing or presenting diagnoses (e.g., anxiety), and/or (b) difficulties related to the first diagnosis present as more pressing clinical concerns (Lai & Baron-Cohen, 2015). This has been observed in a study of children with co-occurring ASD and attention-deficit/hyperactivity disorder (ADHD), where an earlier ADHD diagnosis delayed ASD diagnosis by approximately three years (Miodovnik et al., 2015). However, ASD diagnosis did not significantly delay subsequent ADHD diagnosis (Miodovnik et al., 2015). Coupled with clinician or broader gender biases (explored below), this diagnostic overshadowing may be more problematic for females, if, as evidence suggests, they are more vulnerable to conditions that emerge through internalising behaviours such as depression and anxiety (Oswald et al., 2015; Solomon et al., 2012). Depending on the severity or salience of these other difficulties, they may overshadow ASD difficulties when (and if) the individual comes to clinical attention (Petrou et al., 2018).

In particular, the presence of ASD may be particularly difficult to discern in the context of internalising difficulties that have emerged, perhaps as a consequence of camouflaging. This is because difficulties associated with diagnostic overshadowing may be coupled with behaviours that actively disguise ASD characteristics. As previously stated, it is possible that such individuals may not otherwise come to clinical attention had internalising difficulties or co-occurring conditions not developed. The emergence of such co-occurring psychiatric conditions that are secondary to undiagnosed ASD highlights the importance of early identification. Co-occurring conditions may complicate the identification of ASD difficulties in non-ASD specialist clinical settings, but in the event that ASD concerns are identified, may increase the likelihood of ASD diagnosis (given that additional emotional challenges may be necessary for girls to be diagnosed with ASD compared to boys; Duvekot et al., 2016).

The diagnostic assessment process of females may also be complicated by differences in the severity of externalising difficulties between males and females. Generally, studies examining management of ASD characteristics amongst children have found that autistic girls may be less likely to demonstrate externalising behaviours such as aggression or hyperactivity than autistic boys (Hiller et al., 2014; Mandy et al., 2011). However, some studies have found no significant differences in externalising/internalising behaviours, perhaps as a result of the age of participants or ASD severity (e.g., Nasca et al., 2019; Pisula et al., 2017). If externalising behaviours are greater amongst boys with ASD, females may be less identifiable by professionals such as teachers because their behaviour is less disruptive, and this may delay referral for assessment (Hiller et al., 2014).

Challenges Associated with Current Assessment Methods

Various difficulties associated with the current methods for assessing and diagnosing ASD in females have been identified and may play a role in their possible under-

identification and underdiagnosis of ASD. These include the failure of many assessment instruments to consider the distribution of ASD traits in the typically developing population (greater among boys; Constantino & Todd, 2003) or adequately capture the female ASD presentation. Additionally, concerns have been raised around gender expectations in identifying atypical behaviour.

Assessment Instruments and Typically Developing Comparisons. At present, many of the common assessment instruments used to inform the process of diagnostic assessment (such as the Autism Diagnostic Interview- Revised; ADI-R; Lord, Rutter, & Le Couteur, 1994) do not consider the distribution of ASD traits in the typically developing population, in the context of which abnormality is (or is not) recognised. This is problematic, as, for example, social difficulties in girls with ASD must be considered in the context of the typically superior social skills of typically developing girls compared to typically developing boys (Kreiser & White, 2014; Leman & Tenenbaum, 2011). Similarly, regardless of whether they are autistic, males may demonstrate higher RRBI scores on the Autism Diagnostic Observation Schedule, 2nd Edition (ADOS2; Lord et al., 2000), indicating that sex/gender differences in ASD traits may not be specific to the disorder (Messinger et al., 2015). Thus, given the present single set of ASD criteria for males and females and *non* sex/gender-normed assessment instruments, females may require a greater amount of atypicality (relative to males) in order to meet these thresholds and qualify for ASD diagnosis (Constantino & Charman, 2012; Goldman, 2013). It is for this reason that some researchers have suggested sex/gender-specific ASD criteria and thresholds should be used (Lai et al., 2011).

Assessment Instruments and the Female ASD Presentation. Given the growing evidence supporting the existence of a female ASD presentation, it is possible that the standardised assessment instruments used do not adequately capture the quantitatively and qualitatively distinct presentation (Kreiser & White, 2014; Rutter et al., 2003). As argued

above, the conceptualisation of ASD has been based upon androcentric samples which reflect the historically imbalanced prevalence ratio. Some of the most commonly used assessment instruments have been shown to lack sensitivity to the female phenotype and thus may contribute to the under-detection of ASD in females (Beggiato et al., 2017; Lai et al., 2011).

In their comparison of men and women with ASD, Lai et al. (2011) used two gold-standard diagnostic tools, the ADI-R and ADOS-2, to quantify cognitive and behavioural differences between the sexes/genders. The authors found that a large subset of females who met ASD criteria according to both the ADI-R and the judgement of experienced clinicians, but failed to meet the criteria based on the ADOS-2 (80% of females compared with 43% of males). This was also found in a study of adults who received ASD diagnoses (Adamou et al., 2018) and among a small sample of adolescents (Rynkiewicz & Łucka, 2015), where females scored significantly lower than males on the ADOS-2. Lower ADOS-2 scores on repetitive and restricted behaviours have also been reported in girls compared to boys from analysis of a large sample of autistic children (Kaat et al., 2020). Together, these findings suggest that this instrument may be insensitive to the female phenotype of ASD, perhaps because of its emphasis on clinically observable difficulty which may be attenuated by camouflaging.

Beggiato et al. (2017) recently conducted an investigation into the specific items in the ADI-R which discriminate between males and females with ASD. They found that six items significantly discriminated sex/gender, four of which were included in the diagnostic algorithm. Specifically, girls scored higher than boys (i.e., showed less atypicality) in *The range of facial expressions used to communicate* (Reciprocal Social Interactions domain) and *Imaginative play* (Communication Impairment domain). In contrast, *Circumscribed interests* and *Unusual preoccupations* (Repetitive and Stereotyped Behaviours domain) were more pronounced among boys. In order to avoid sex/gender bias when using the ADI-R, the authors suggest using correction factors on these items. However, it is not currently clear how

the algorithm can be adjusted to be sex/gender-neutral, or how these corrections might be applied. In light of this evidence, although the use of these instruments constitutes a methodological strength in the broader literature, they may poorly capture the female ASD phenotype and therefore underestimate the challenges of females.

As argued by van Wijngaarden-Cremers et al. (2014), only females who present with behaviours consistent with the male presentation are likely to be captured by assessment instruments (such as the ADOS-2) and thus diagnosed with ASD (Kirkovski et al., 2013). A serious limitation of the current literature on sex/gender differences in ASD and reasons for underdiagnosis in females is that only individuals diagnosed with ASD (and embody the typical male presentation) are usually included (Lai & Baron-Cohen, 2015). As a result, very little is known about the experiences and behavioural presentation of females for whom ASD is suspected, but not diagnosed. It is therefore important that future research consider the presentation of females who are determined to be narrowly below the diagnostic threshold for ASD.

The Role of Gender in the Identification of ASD. Kreiser and White (2014) argue that culture-based gender expectations of typical child development may also contribute to the proposed under-identification of ASD among females. The authors suggest that this occurs in two ways. First, gender expectations provide a context within which abnormality is or is not perceived (e.g., social withdrawal or unusual behaviour may be perceived as ‘shyness’ or ‘immaturity’ instead of indicative of ASD; Attwood et al., 2006). Second, societal influences may moderate the expression of ASD difficulties depending upon the child’s gender through their interactions with other people, including differential reactions to norm violations, and the reinforcement and shaping of behaviour (Cheslack-Postava & Jordan-Young, 2012). For example, the expectation that girls should be social and demonstrate friendliness and empathy may drive parents and teachers to encourage social

behaviour more among girls with ASD than in boys with ASD (Kreiser & White, 2014; Rivet & Matson, 2011b). Indeed, Kreiser and White (2014) suggest that the influence of reinforcement and punishment for norm-violation may be greater for females than males in many cultures. It is possible that these culturally based gender expectations may account for at least some of the aforementioned differences in ASD behavioural expression between males and females, and why ASD-related abnormality may not be as identifiable in females. The lack of sex/gender differences among toddlers (van Wijngaarden-Cremers et al., 2014) may be partially explained by gender socialisation, which may occur over time. Equally, it is possible that girls who do show atypical behaviours may be viewed as more deviant and therefore more likely to be referred for assessment, whereas in boys, this may be perceived as less abnormal. The literature has thus far been unable to reconcile these conflicting perspectives.

Gender Expectancy Bias. A final hypothesis surrounding the possible underdiagnosis of females with ASD pertains to the expectations of referring parties, such as teachers, parents, and clinicians screening and conducting formal assessments for ASD. According to the gender expectancy bias hypothesis, referrers and clinicians may be less likely to consider, raise concern about, or diagnose a particular disorder if it is (a) less common in one sex/gender (in this case females), or (b) when features of the disorder are considered more typical of one sex/gender (Hartung & Widiger, 1998). This effect has been demonstrated in studies examining clinicians' decision-making in the diagnosis of depression, where a false negative diagnosis is more common in men (Potts et al., 1991), and certain personality disorders (e.g., histrionic and borderline personality disorder), where diagnosis is more common in women (Worell & Robinson, 2009). In ASD, it is possible that, purely by virtue of the disorder being diagnosed more often in males, it is looked for more often in males. Similarly, in keeping with the Extreme Male Brain theory (i.e., that male

brains have a propensity for systemising over empathising; Baron-Cohen, 2002) and the unequal distribution of ASD traits in the typically developing population (Constantino & Todd, 2003), the possibility of ASD may be more forthcoming for males.

Recent experimental evidence suggests that a child's sex/gender influences referrers' perceptions of their ASD characteristics. Geelhand et al. (2019) provided members of the general public with identical descriptions of behaviours commonly seen in ASD, randomly assigning the sex/gender of the child described to each vignette. While sex/gender did not significantly affect participants' levels of concern, females were significantly less likely than males to be thought to demonstrate future atypicality in adolescence (i.e., more likely to 'grow out of' these behaviours). Similarly, bias has also been found among educators presented with a series of vignettes (including a typical 'male' and 'female' ASD presentation), in which the sex/gender of the child was randomly allocated across vignettes and participants (Whitlock et al., 2020). Here, educators were more likely to correctly identify ASD when the vignettes described males than when they described females. Furthermore, educators were less able to identify autism in the female ASD presentation vignettes compared to the classic male presentation. Together, these studies provide preliminary evidence to suggest that, for children presenting with identical autistic behaviours, female sex/gender may reduce the likelihood that concern will be raised.

Somewhat alarmingly, a small number of investigations has shown that in the event that a girl is referred for specialist assessment, she may be less likely to receive an ASD diagnosis than a male with ASD difficulties of identical severity (Dworzynski et al., 2012; Giarelli et al., 2010; Russell et al., 2011). However, none have taken an experimental design to explore alternative diagnoses, perceived severity levels and in which domains (if any) females are less likely to display clinically significant difficulties. To date, diagnosticians' voices have contributed little to the literature, and their perspectives of sex/gender

differences, challenges assessing females for ASD, and means for circumventing these challenges remain unknown. Only two studies have directly investigated diagnosticians' experiences, identifying whether they may conceptualise ASD slightly differently in females. Specifically, diagnosticians have noted more sex/gender differences in RRBIs than in social domains (Jamison et al., 2018), and that girls may manage the condition differently which impacts upon the expression of difficulties (Muggleton et al., 2019).

In qualitative investigations examining the females' diagnostic experiences, autistic adolescent girls, women, and their families have raised concern that there is a lack of understanding among professionals regarding the female presentation. According to these women (Baldwin & Costley, 2016), adolescent girls (Bargiela et al., 2016; Cridland et al., 2014), and their parents (Navot et al., 2017; Rabbitte et al., 2017), this has led to delays in diagnosis, misdiagnosis, and/or scepticism as to whether a 'problem' truly exists. In turn, this resulted in the girls and women with ASD feeling misunderstood, unfairly labelled, and unsupported or inadequately supported at school and in other contexts (Bargiela et al., 2016; Cridland et al., 2014; Navot et al., 2017). Such uncertainty, distress, and inadequate support heightens symptoms of anxiety and depression and risk of suicidality in these individuals (Cassidy et al., 2018). Consequences such as these highlight the critical importance of ensuring both the timely and accurate identification and diagnosis of ASD in female clients.

Structure and Contribution of the Thesis

This thesis makes a number of contributions to the existing literature on sex/gender differences in the presentation of ASD and reasons for the apparent underdiagnosis of females. Bayesian statistical analyses are applied throughout and the advantages of this approach are discussed in Chapter 2: Statistical Analysis. In Study 1 (Chapter 3), I explored sex/gender differences in ASD presentation in a large number of autistic children according to two established ASD screening instruments. Rather than relying on a small number of

summary scores, this was done at item level in order to shed light upon differences in the severity of specific behaviours (i.e., qualitative differences in presentation) and the likelihood that they would emerge as *clinically significant*. The sensitivity of these instruments to the difficulties posed by females in the diagnostic process is discussed.

Another important contribution of this thesis is the examination of ASD characteristics of children narrowly below the ASD diagnostic threshold (i.e., those found not to fully meet ASD criteria despite having many ASD traits, being suspected of having ASD and having undertaken formal developmental assessment; Study 2, Parts A and B; Chapter 4). The differences between the two presentations of children who returned for follow-up assessment (at which they were diagnosed with ASD; Study 2a) were considered separately to enable longitudinal analysis of emerging ASD difficulties. In examining the difficulties of females under the diagnostic threshold, I incorporated fine-grained diagnostic information gathered via parent interview, the diagnosticians' clinical observations, and feedback from teachers. Thus, possible differences in ASD presentations based on the social environment are examined. This study thus allowed for identification of where females may fall short of fulfilling diagnostic criteria, why this might be the case, and how the ASD conceptualisation could be modified to better capture their specific difficulties.

Finally, in Study 3 (Chapter 5) I explored diagnosticians' experiences in assessing females for ASD and their perspectives as to why the condition may be more difficult to identify among females. I also adopted an experimental design to examine potential sex/gender-related biases in diagnostic decision-making. Broadly, findings from this thesis assist in better understanding the female presentation of ASD and why it may be under-detected.

Chapter 2: Statistical Analysis

The major statistical analyses presented in this thesis have been performed using a Bayesian parameter estimation approach. Bayesian analyses offer a number of advantages over null-hypothesis significance testing (often referred to as frequentist or classical statistics) approaches. These advantages are numerous and have been discussed at length by statisticians over the past 40 years or more (for a review of these advantages, see Kruschke, 2010). Some of these advantages are particularly pertinent to this thesis. Most importantly, Bayesian analyses provide more direct and useful estimates of the information scientists need to answer data-based questions. Also, Bayesian analyses are flexibly adapted to hierarchical data structures (e.g., multiple questions answered by a single participant or multiple assessments of the same client) in a way that improves the quality of parameter estimates (Wagenmakers et al., 2017). In contrast to classical analyses, Bayesian approaches manage well with small samples. These advantages stem from a number of core differences in the underlying nature of Bayesian and classical analyses. These differences and their importance in this thesis are discussed below.

Prior and Posterior Parameters

A Bayesian approach begins with *prior distributions*, or a description of what is already known about the parameters of a particular model. Ultimately therefore, the results produced (posterior distribution for each parameter) are influenced by the pre-existing knowledge relating to the research question. In this way, Bayesian analyses allow researchers to statistically update existing knowledge on a subject using the new data. Importantly, this description of knowledge can reflect a well-understood phenomenon (entailing relatively precise parameter estimates from previous data) through to complete naivety (entailing no information beyond the practical limits of values). The overarching research questions presented in this thesis, while not all necessarily novel, have not been tested using the same

methodology or investigated using a Bayesian approach. For the models presented in the results of this thesis, I have therefore used noncommittal priors that reflect the lack of current knowledge surrounding possible values for the model parameters. This means that the results reflect the data without the influence of pre-existing knowledge. Thus, the approach in this thesis is comparable with classical analyses that (with the exception of meta-analyses) incorporate no information from previous studies (Kruschke, 2014).

More Informative than Null Hypothesis Significance Testing

Compared to frequentist approaches, Bayesian analyses are more in keeping with the New Statistics movement (Cumming, 2012), which calls for more informative illustrations of effect sizes, paired with a description of the uncertainty around these effect sizes (Kruschke & Liddell, 2018). This movement cautions against null hypothesis significance testing (NHST) which only supports two conclusions: (a) the null hypothesis can be rejected, and (b) the null hypothesis cannot be rejected. Importantly, standard NHST does not support the acceptance of any specific alternative hypothesis, nor does it allow acceptance of the null. In contrast, Bayesian analyses not only support all of these conclusions, but also provide quantitative information about how likely the conclusions are, given the data. Bayesian analyses more naturally provide researchers with the information they seek (Wagenmakers et al., 2017): given the data collected, what are the most credible statistical parameters and how confident can we be in these values?

Frequentist analyses produce point estimates (or a single ‘best guess’) for statistical parameters (e.g., mean, standard deviation, or effect size). In contrast, Bayesian models produce distributions that reflect the most plausible values for a parameter given the data and, importantly, illustrate the credibility of the proposed parameters. Although frequentist analyses can produce confidence intervals which present a range, these intervals are difficult to interpret and do not provide the information that many researchers mistakenly infer

(Morey et al., 2015). For example, the width of confidence intervals do not quantify the precision of an estimate, confidence intervals are not a sound guide as to which parameter values are plausible (or implausible), and confidence intervals cannot support any probabilistic conclusion (i.e., we cannot conclude that there is an $x\%$ probability that the true parameter is contained in a $y\%$ confidence interval). In contrast, Bayesian posterior distributions can be used to calculate credibility intervals. These credibility intervals do, in fact, support the conclusions listed above that are mistakenly made on the basis of confidence intervals.[9]

This thesis uses highest density intervals (HDIs; a special case of the Bayesian credibility interval). These intervals contain the range of values that have a specific probability of containing the true value and include the most credible parameter values (i.e., no value outside the HDI is more credible than a value within the HDI). In classical analyses, the arbitrary level of 95% is typically used. As Bayesian analyses are sensitive to more sources of uncertainty in parameter estimates than classical analyses (Gelman et al., 2014), an unthinking use of 95% is not always appropriate, especially in exploratory work. Consequently, this thesis uses a criterion of 80%.

The $\text{HDI}_{80\%}$ captures the range of values which are the most credible and cover 80% of the distribution of possible parameters. One can have 80% certainty that the true value of a parameter lies within the specified range. Therefore, the width of the HDI (or the spread of values within it) illustrates the level of certainty of the model for a particular parameter.

[9] Interestingly, the only situations in which confidence intervals can be used as a basis for the listed conclusions is when they are shown to agree with a Bayesian credibility interval. As Morey et al. (2015) note, given that the only way to usefully interpret a confidence interval (for the vast majority of questions posed by researchers) is to show that it matches the Bayesian credibility interval, simply using the Bayesian credibility interval seems a more appropriate first step.

Defining a Null Result

Another important distinction between frequentist and Bayesian parameter estimation approaches is how the criterion for *no difference* between groups, or a null result, is established. In frequentist analyses, alternate hypotheses are tested against the null which states that there is a difference of zero between groups. Depending on the strength of the evidence collected, one then concludes whether or not the null hypothesis (or difference of zero) can be rejected. However, it is problematic to compare to exactly zero as the criterion for no meaningful effect (Kruschke & Liddell, 2018; Kruschke & Vanpaemel, 2015). This is because a researcher is unlikely to be interested in whether the difference between groups is not exactly zero, but rather whether the difference is large enough to be of interest or meaningful. Further, comparison with a point null of zero creates a logical problem (Lindley's paradox) whereby increasingly precise estimates (e.g., with bigger samples and/or better measurement) are increasingly likely to reject the null regardless of its truth. In other words, paradoxically, increasingly better data lead to increasing probability of an erroneous conclusion when the null is in fact true.

The Bayesian region of practical equivalence (ROPE) approach allows us to define an interval within which all values are of practical equivalence to the null hypothesis and indicate a non-meaningful (or negligible) difference, but not necessarily a difference of exactly zero. The following three scenarios illustrate the conclusions that can be drawn from the relationship between the $HDI_{80\%}$ and ROPE:

1. If the $HDI_{80\%}$ lies completely outside the ROPE, we can have 80% confidence that there is a meaningful effect in this direction, where *meaningful* is defined by the ROPE criterion.
2. If the $HDI_{80\%}$ falls completely within the ROPE, we can conclude, with 80% confidence, that there is evidence of no meaningful effect or relationship.

3. If the $HDI_{80\%}$ *partially* overlaps the ROPE, the evidence is equivocal (at the 80% confidence level) regarding accepting versus rejecting the null. However, in contrast with classical analyses, further examination of the posterior distribution can quantify the extent to which the balance of evidence favours the null or the hypothesis. Specifically, the proportion of the posterior distribution lying outside (or within) the ROPE quantifies the probability of a meaningful (or null) effect.[10]

Thus, ROPE-based conclusions favour a difference when the data clearly shows that the effect is meaningfully sized, not just non-zero. Importantly, this result when in the predicted direction, provides direct support for the hypothesis (in contrast, NHST results than can only ever reject the null, but never directly speak to the extent to which the data support the hypothesis). Furthermore, increasingly precise estimates lead to an increased probability of accepting the null when it is true and of accepting the alternative when there is, in fact, an effect. In other words, the ROPE approach leads to increased probability of an accurate conclusion with increasingly strong data (regardless of the truth/falsity of the null).

The first two scenarios presented above arise when the data clearly support a conclusion (either for the hypothesis or for the null) at the given level of certainty (80% in this thesis), and thus support the strongest conclusions regarding the presence or absence of a difference. In this thesis, I use the term *meaningful effect* to refer to results that support this type of clear conclusion; it can be thought of as an analogue of the classical term *significant effect*.[11] Additionally, the term *evidence of no effect* will be used to refer to a conclusion in favour of the null (a conclusion that cannot be supported by classical analyses and, therefore,

[10] Refer to page Figure 2.1 (page 43) for a worked example of interpreting HDIs relative to the ROPE.

[11] As already discussed, the evidence provided by this type of analysis speaks directly to the level of support for the hypothesis. In contrast, NHST analyses only quantify the consistency of the data with the null; they do not speak directly to the plausibility or level of support for the hypothesis itself.

has no common NHST analogue). In this thesis, when the evidence is not strong enough to support either of these clear conclusions, the proportion of the posterior that lies outside (within) the ROPE will be reported to quantify the balance of evidence in favour of the hypothesis (null). These probabilities can help distinguish between situations where the evidence is genuinely equivocal, versus favouring an effect (or the null) but with insufficient certainty to support the conclusion of a meaningful effect. The probability that the true difference falls outside the ROPE in the direction of the observed effect is denoted $P_{(meaningful)}$, and the probability that it falls within the ROPE (negligible difference) is denoted $P_{(within\ ROPE)}$.

Mixed Modelling

A final advantage of Bayesian parameter estimation approaches is that they more easily and flexibly manage hierarchical or mixed-effects modelling in analysing the results of complex methodological designs (Gelman et al., 2014; Kruschke & Vanpaemel, 2015). Hierarchical models are applied in instances where it is important to model the average of a group and to account for variance in these estimates due to variation between individuals in the sample and variance between the items on which they are assessed (or the questions they answer). By explicitly modelling the variance of parameters between participants and between items, the analysis provides a superior estimate of group-level performance and of the differences between groups (Gelman et al., 2014; Gelman et al., 2012). Importantly, when conducted on multiple items, such analyses can then be used to examine overall patterns aggregated across items as well as to explore group differences at an item-by-item level. Importantly, such analyses provide superior estimates than either analysing aggregated data or individually analysing each item (Gelman & Hill, 2006).

Application in This Thesis

The output of Bayesian analyses is a joint probability distribution of all model parameters. From these, relevant statistics can be calculated. In this thesis, aggregate statistics (means and proportions, for numeric and discrete variables, respectively) are used as descriptive statistics for each group, and effect sizes for contrasts between groups (Cohen's *d* for numeric variables and log odds ratios for discrete variables). Across each study, male sex/gender is coded such that a positive effect size reflects a higher score for males (either mean or proportion; M > F is reflected by a positive effect size), and female sex/gender is assigned the negative direction (i.e., F > M is reflected by a negative effect size).

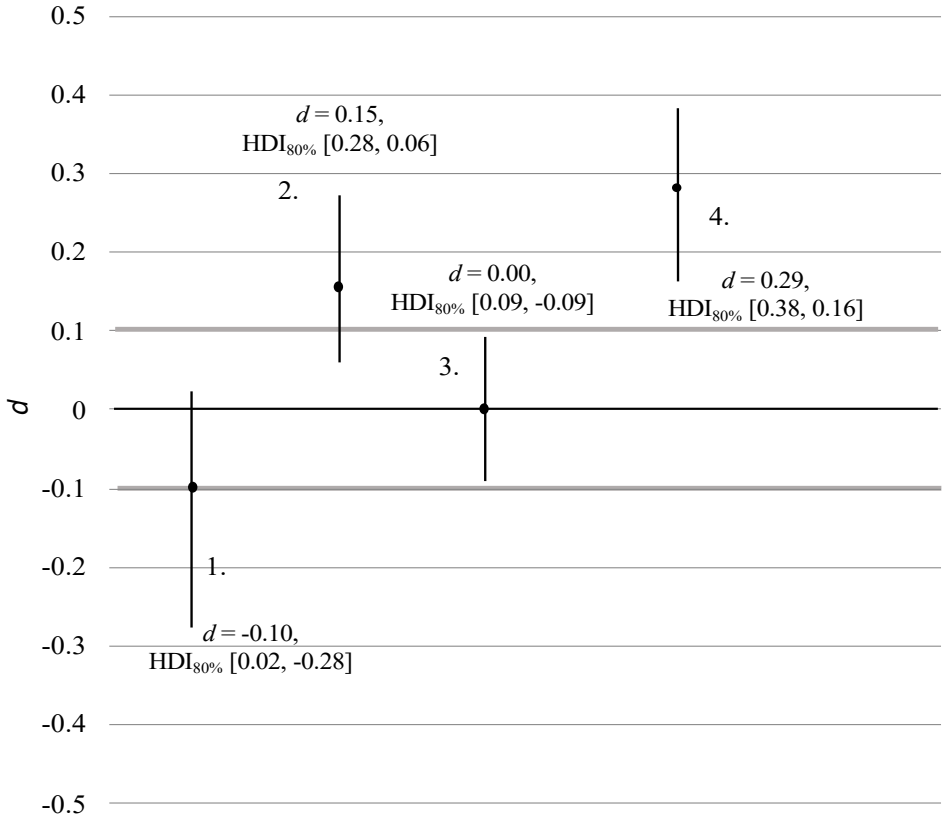
To allow for consistent interpretation across studies, I set the ROPE on the effect sizes, rather than scale-dependent indices (e.g., unstandardised mean differences). For numeric comparisons using Cohen's *d*, the ROPE was [-0.1, 0.1]. This was based on Kruschke's (2018) recommendation to use half the typical cut-off for a small effect in situations where domain/measure specific knowledge does not suggest a more informed definition of a trivial or unimportant effect size. For discrete variable comparisons using log odds ratio (LOR), the ROPE was defined as [-0.1, 0.1] (again, following Kruschke's recommendations for determining a ROPE).[12]

R software (R Core Team, 2019) with JAGS (Plummer, 2017) and the runjags package (Denwood, 2016) were used for all statistical analyses. Results of the major analyses are displayed in plots, such as the example in Figure 2.1, which illustrate the distribution of posterior values resulting from the models. Figures were produced using the ggplot (Wickham, 2009) and cowplot (Wilke, 2016) packages. Estimates reported in text or table (e.g., mean or effect size) are reported with an $HDI_{80\%}$ [upper, lower bounds].

[12] The adoption of the same ROPE for both *d* and LOR was a coincidence.

Figure 2.1.

Example Plot of Effect Size Estimates



Note. Error bars represent the HDIs $_{(80\%)}$ for each of the four cases. Horizontal grey lines depict the ROPE's upper and lower bounds (± 0.1). Differences are judged to be meaningful when the $HDI_{80\%}$ is entirely outside of the ROPE. Differences in the positive direction indicate an effect in the direction of group A, and differences in the negative direction indicate an effect in the direction of group B.

The interpretation for each of the four numbered effect size estimates displayed in Figure 2.1 will be presented in turn:

1. Here, the $HDI_{80\%}$ overlaps the ROPE. Evidence of a difference between the groups is therefore equivocal: neither a negligible difference nor a difference in either direction can be accepted or rejected with 80% confidence. Approximately 60% of the posterior distribution falls outside the ROPE and therefore there is a 60% probability that the true difference is meaningful and in the negative direction ($P_{(meaningful)}$ = 60%, $P_{(within}$

$_{ROPE)}$ = 40%). Thus, the evidence slightly favours a difference, but too weakly to support any conclusion.

2. As most of the $HDI_{80\%}$ falls outside the ROPE, a probable difference in the positive direction can be inferred. However, as a portion of the $HDI_{80\%}$ lies within the ROPE, a negligible difference cannot be excluded with 80% confidence. Thus, the evidence does not support the conclusion of a meaningful difference (at the 80% confidence level). Seventy-nine percent of the posterior distribution lies above $d = 0.1$ (the upper bound of the ROPE) and therefore $P_{(meaningful)}$ = 79%. This means that there is sufficient evidence to support a weak conclusion of a (non-negligibly sized) effect in the positive direction.
3. The $HDI_{80\%}$ falls entirely within the ROPE, therefore we can conclude, with 80% confidence, that there is no meaningful difference between groups.
4. The $HDI_{80\%}$ is entirely above the ROPE and therefore we can conclude, with 80% confidence, that the difference is meaningful and in the positive direction.

Chapter 3: Study 1

Sex/Gender Differences in CARS-2 and GARS-3 Item Scores: Evidence of Phenotypic Differences Between Males and Females with ASD

Overview

When reflecting upon autism spectrum disorder (ASD), a group of unusual behaviours and an image of a *male* child will for many, come to mind. This stereotype may have originated with Kanner and Asperger's seminal case studies (1944; 1943), the large majority of which described the difficulties of male patients. It is now understood females may also present with ASD, but perhaps as a result of our historical understanding of the condition and the hypotheses posed in Chapter 1, they may be less likely than males to receive a diagnosis. While lower prevalence of ASD among females may be due to biological protective factors, growing evidence suggests that females who do develop ASD are more likely than males to be diagnostically overlooked. An important hypothesis attempting to account for this phenomenon posits that subtle quantitative and qualitative differences exist between the typical ASD presentations of males and females (Lai et al., 2011; van Wijngaarden-Cremers et al., 2014). A large part of the argument and rationale for the present study is foreshadowed in Chapter 1. For the sake of brevity, only the issues of direct relevance to the present study will be outlined here. As argued in Chapter 1, where females' ASD symptoms deviate from the well-established androcentric conceptualisation and are imperfectly captured by instruments used to detect ASD, symptoms may be less likely to be identified as autistic (Lai, Baron-Cohen, et al., 2015). Ultimately, this results in under-diagnosis. Although existing diagnostic instruments may perpetuate stereotypical 'male' notions of ASD, analysis of sex/gender differences in symptom profiles may assist in better understanding the female presentation where it deviates from the male presentation and whether females are therefore less likely to reach diagnostic thresholds. To this end, the

present study will examine sex/gender differences in symptom profiles on The Childhood Autism Rating Scale, 2nd Edition (CARS2 Standard and High Functioning forms; Scholper et al., 2010), one of the most frequently used instruments in ASD assessment. This study will also examine fine-grained sex/gender differences using the Gilliam Autism Rating Scale-3 (GARS-3; Gilliam, 2014), in order to clarify how sex/gender may influence the expression of specific ASD features.

Historically, males with ASD have outnumbered females across all stages of development. While current estimates suggest that four to five males are diagnosed with the condition for every female, evidence from a large population study indicates a smaller difference between males and females in *true incidence* (3 males: 1 female; Loomes et al., 2017). Although biogenetic factors may mean that males are more vulnerable to developing ASD, growing evidence suggests that a large number of females with ASD remain undiagnosed. This may be because females are less likely to be referred for specialist assessment than males, and in the event that they are considered for diagnosis (Hull et al., 2020), may need to display greater ASD or associated difficulties (e.g., emotional or intellectual) in order for an ASD diagnosis to be provided (Duvekot et al., 2016; Dworzynski et al., 2012; Ratto et al., 2018).

Two broad but related hypotheses have been put forward to explain this apparent underdiagnosis. Perhaps as a result of the prevalence imbalance and historical assumption that ASD is primarily a 'male' disorder, clinicians and referrers may be less inclined to consider ASD or pursue further investigation for female patients (Russell et al., 2011). Compounding this bias, growing evidence suggests that subtle differences in ASD presentation may exist between males and females (Hull et al., 2020; Kirkovski et al., 2013). The female presentation of ASD is not yet well understood, and in particular, it remains unclear whether females typically exhibit a milder or more functional presentation than

males, whether females differ in how they manage or conceal their symptoms, and/or whether their symptoms differ in expression or manifestation, rendering symptoms less readily detected.

Quantitative Phenotypic Differences

Studies comparing the severity of social and communicative difficulties have produced inconsistent results. Specifically, they have identified greater (Carter et al., 2007) and lesser atypicality amongst females (Lai et al., 2011; McLennan et al., 1993), and no significant difference in severity between males and females (Mandy et al., 2011). Taken together in Wijngaarden-Cremers and colleagues' meta-analysis of sex/gender differences in core ASD features (2014), autistic males and females were found to be comparable in their levels of social difficulties (std. mean diff. = -0.04, $CI_{95\%}$ = [-0.20, 0.13], p = 0.66) and communicative difficulties (std. mean diff. = -0.03, $CI_{95\%}$ = [-0.26, 0.21], p = 0.82). However, repetitive and restricted behaviours and interests (RRBIs) have been found to be less pronounced among females in many studies (Allely, 2019). As all children recruited in the above studies had ASD diagnoses and thus showed enough broad ASD-related behaviour to meet criteria, it is perhaps unsurprising that only small differences in severity have been found. In other words, inclusion of those who are mistakenly overlooked for ASD diagnosis could reveal a larger overall sex/gender difference.

In addition to variables such as age and cognitive ability (Hull et al., 2016; van Wijngaarden-Cremers et al., 2014), differences in how males and females typically manage their ASD symptoms may affect their outward expression and others' impressions of symptom severity. The development of behaviours used to manage ASD traits over time may help explain the influence of age and intellectual ability on observed sex/gender differences. In particular, recent research has focused on camouflaging, or the constellation of behaviours by which an autistic individual strives to conceal and compensate for their ASD features in

particular settings (Hull et al., 2017; Lai et al., 2016). Camouflaging is currently considered an important feature of the female ASD phenotype, as females may be better able (Lehnhardt et al., 2016) and/or more motivated to camouflage; perhaps as a result of cognitive factors or socially reinforced gender expectations (Hull et al., 2017). As a result of the concealment and compensation of autistic features and internalisation rather than externalisation of distress (Hiller et al., 2014; Solomon et al., 2012), females' difficulties may be mislabelled, often as part of an anxiety or mood disorder (Bargiela et al., 2016). Furthermore, if females are less likely to outwardly display their difficulties or present with disruptive behaviours, their difficulties may escape the attention of parents, teachers, and even healthcare professionals.

Qualitative Phenotypic Differences

Differences between males and females in the management of ASD symptoms imply that, while they may or may not differ in overall severity, symptoms may differ qualitatively between males and females (i.e., in the manifestation of broad ASD-related difficulties into specific behaviours). This would result in sex/gender-specific behavioural profiles. Lai et al. (2015) propose a theoretical framework for conceptualising possible sex/gender differences in ASD presentation. They suggest that differences may be least apparent at the level of the broad domains that define ASD (i.e., the severity of social communication and RRBI difficulties), but that phenotypic differences may be most apparent at the level of the "fine-grained subdomains" (p. 13) which comprise the above, and specific behaviours which lie within each subdomain. If this hypothesis is borne out, then sex/gender differences are unlikely to be detected by examining overall scores on diagnostic instruments, as has been the focus of many investigations to date.

The small number of studies comparing the presentations of autistic males and females at the level of specific behaviours has indeed found differences within the broader social communication and RRBI domains. For example, Hiller et al. (2014) found that girls

may be less impaired than boys in their ability to share interests, engage in social and emotional reciprocity, and to initiate friendships, but not significantly different in their ability to maintain friendships. Similarly, although females may demonstrate slightly less severe overall RRBI characteristics, they may be more likely to engage in specific stereotypical behaviours such as self-injury, whereas males may be more likely to demonstrate stereotypical movements and unusual preoccupations or restricted interests (Antezana et al., 2018; Beggiato et al., 2017).

If, as suggested above, some females present in a way that deviates from the 'classic' male ASD presentation based on which the disorder was conceptualised, their symptoms may be less likely to be recognised as indicative of ASD (Kirkovski et al., 2013; Mandy et al., 2011). As a result, these females may be less likely to present for ASD assessment or to meet the existing criteria in the event that they do present.

Diagnostic Instruments

As ASD screening and diagnostic instruments were designed to assess for the presence and severity of ASD based on androcentric literature, they may lack sensitivity to the difficulties of some females, particularly where their presentation deviates from the classic ASD presentation (Lai & Baron-Cohen, 2015; Rivet & Matson, 2011b). However, such instruments may be useful in identifying areas in which females' presentations deviate from males' (i.e., in which areas females demonstrate higher or lower levels of atypicality) if their profiles are considered at the level of subdomains and specific behaviours. Additionally, it may be informative to explore specific areas in which females are more or less likely than males to present as typically developing (i.e., with an *absence* of impairment). In doing so, we may be able to clarify if and to what extent current instruments lack sensitivity to the ASD female presentation and how they may best be modified.

The few studies examining sex/gender differences in narrow ASD subdomains and behaviour profiles have commonly employed instruments which rely on single sources of information (e.g., parent report or clinical observation). Given that it is the integration of all available information which enables a clinician to form a diagnostic opinion, analyses of sex/gender differences in behavioural subdomains using integrative instruments may be useful in better understanding the female ASD presentation and reasons for its under-detection. The Childhood Autism Rating Scale, 2nd Edition (CARS2; Scholper et al., 2010) is one of few instruments which is frequently used, allows integration of all diagnostic information, and produces individual scores for a series of behavioural clusters, with scope for the clinician's general impression of the child's ASD severity. Each item is scored according to the presence of ASD features, the level of abnormality with which a child presents (compared to that observed in typical development), and the degree to which the abnormality interferes with daily functioning. A version of the CARS has been used to examine sex/gender differences in ASD behavioural subdomains. Kumazaki et al. (2015) compared item-level scores on the CARS-Tokyo version (CARS-TV; Scholper et al., 1988; similar to the current CARS2-ST) in 46 children (n = 20 females) aged between five and nine years with high-functioning ASD.[13] While overall ASD severity did not differ significantly according to sex/gender, the authors reported greater atypicality in repetitive and stereotyped behaviours and hyperactivity amongst males, and greater anxiety amongst females, consistent with previous literature (for a recent review, see Hull et al., 2020). Additionally, females

[13] The term 'high-functioning ASD' is problematic and may be misleading. It has frequently been used to refer to autistic individuals with higher cognitive ability, milder symptoms and better long-term outcomes. However, many have argued that appraisals of 'functioning' levels should be based upon the (non-stable) adaptive functioning capacity of the individual, which is only weakly related to cognitive ability (Alvares et al., 2020). In this chapter, the term 'high-functioning' is used for consistency with the terminology of the CARS2-HF form, with awareness of its limitations.

were found to have greater atypicality in *Taste, smell, and touch response and use.* The mixture of results consistent and inconsistent with the broader literature is perhaps unsurprising due to the small sample size. Here, I extend the work of Kumazaki et al. by examining the extent to which males and females, in a large sample of Australian children, differ in ASD symptom profiles as depicted by the CARS2.

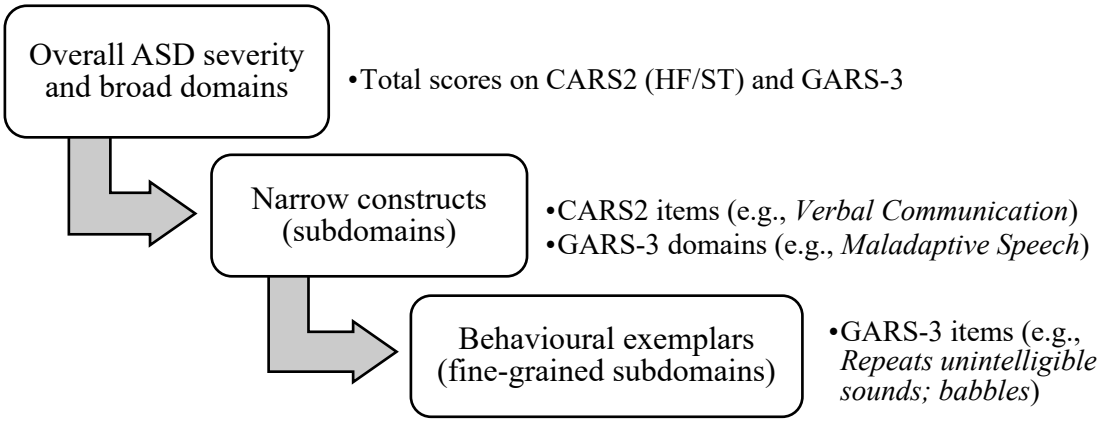
While the CARS2 provides measures of ASD characteristics in behaviour clusters (subdomains) as an aggregate of all diagnostic information, it does not give fine-grained detail about specific behaviours. Therefore, to analyse phenotypic sex/gender differences at this level, I also examined children's item-level profiles using another common ASD screening instrument, the Gilliam Autism Rating Scale-3 (GARS-3; Gilliam, 2014). The CARS2 and GARS-3 differ both in how ASD characteristics are clustered and the specificity with which individual ASD-related behaviours are measured. Although the GARS-3 can be completed by teachers, parents, or clinicians, only parent report forms were available for analysis in this study. Consideration of specific, parent-reported ASD behaviours may be particularly useful in the analysis of sex/gender differences if, as has been suggested, females demonstrate less observable atypicality in structured environments such as during assessment and/or at school (Hiller et al., 2014). Therefore, both sets of profiles were analysed for item level sex/gender differences and the implications of these differences on the detection of ASD in females were considered. Figure 3.1 presents each level of the theoretical framework proposed by Lai et al. (2015) and illustrates where each of the domains and items of the above instruments fall within this framework. The research questions were as follows:

1. In which specific ASD behaviours and behaviour subdomains do females (or males) with ASD differ in the severity of atypicality demonstrated?
2. In which specific behaviours are females (or males) with ASD less likely to present with any atypicality suggestive of ASD?

Given the exploratory nature of this investigation, no formal hypotheses were established.

Figure 3.1

Theoretical Framework (Adapted from Lai, Lombardo, et al., 2015)



Method

Participants

Data were collected from the ASD diagnostic assessment reports of 523 male and 255 female children (total $N = 777$) who were clients of a large private clinic specialising in assessment and intervention for pervasive developmental disorders. Data from both dual assessments (i.e., conducted by a psychologist and speech pathologist) and single assessments (i.e., conducted by only one professional) were included in the current study. Assessors included seven speech pathologists and four psychologists, all of whom have undertaken training for ASD assessment and are recognised as diagnosticians by the local Autism organisation. As per the clinic's standard intake procedure, parents or guardians of clients provided informed consent for the use of their child's diagnostic information for research purposes. Ethics approval was obtained through the author's tertiary institution.

Procedure

All diagnostic reports of children diagnosed with ASD at this clinic between September 2013 and October 2018 were examined.[14] Prior to (and following) data collection, assessment reports were stored securely in portable document format (PDF) on the clinic's online database, accessible only to employees. Resultantly, data integrity was maintained (i.e., there was no possibility of tampering with scores following the production of profiles at the time of assessment). Children whose diagnostic assessments included the use of a CARS2 and/or GARS-3 were identified (87.6% of children: 89.2% of males and 84.4% of females). The reports of these assessments were categorised according to the instrument used and domain-level and total scores were collected for each. Item-level GARS-3 scores were collected if available (70.3% of children who were scored using the GARS-3; 68.7% of males, 73.0% of females).[15] For the purpose of describing the sample, standardised intellectual ability data were also extracted if available. To minimise the likelihood of human error, all scores were checked immediately after extraction from the assessment reports.

Measures

Childhood Autism Rating Scale- 2nd Edition (CARS2)

The CARS2 is an empirically validated and widely-used instrument which provides quantified summary information about a child's behaviours in specific domains commonly affected by ASD (Scholper et al., 2010). The CARS2 is used as a means to integrate

---

[14] This window was selected because all assessments during this time were conducted under the DSM-5 criteria and according to procedures endorsed by Autism SA (the local autism association). Given that data were extracted from assessment reports within approximately five years of the time at which the scores were generated (under the same guidelines), data latency was not anticipated to meaningfully affect the quality of the data.

[15] See Appendix B (Table B.1) for descriptive statistics on this group. There were no meaningful differences between age or intellectual ability between the groups of children for whom only GARS-3 domain scores were available and those for whom item scores were also available.

information from different sources, inform assessment feedback, and guide support program planning. The instrument includes the Questionnaire for Parents and Caregivers (CARS2-QPC) and two forms for clinicians (Standard, ST and High-Functioning, HF). The latter two forms are completed by professionals trained in ASD assessment and are selected according to the age, intellectual ability, and adaptive functioning level of the child. Scores on these two forms were relevant to the present study.

The CARS2-ST is designed for use with children with an estimated Intelligence Quotient (IQ) less than or equal to 79, with impaired communication, or under the age of six years. The CARS2-HF is designed for children over the age of six years and with an estimated IQ equal to or greater than 80. Although many of the domains are common across the ST and HF forms, the descriptors and operationalisation of severity levels are tailored for use with the corresponding group of children. They are therefore considered separately in this investigation. On both forms, the frequency, atypicality, and intensity of behaviours grouped within each item are scored together from 1.0 (typical) to 4.0 (severely atypical) in increments of 0.5 (item descriptions are presented in Tables 3.2 and 3.3). Summing the scores for each of these 15 items produces a total CARS2 score, which can range from 15.0 (least severe) to 60.0 (most severe). A child's presentation can then be assigned a categorical ASD severity rating (i.e., *Minimal-to-No Symptoms of ASD*, *Mild-to-Moderate Symptoms of ASD,* or *Severe Symptoms of ASD*) according to their total score and age. Scores corresponding with each of these categories differ according to the form used.

Diagnosticians completed the CARS2 forms following behavioural observation of the child and the collection of parent interview information (as well as teacher report data, if available).

Gilliam Autism Rating Scale- 3rd Edition (GARS-3)

The GARS-3 (Gilliam, 2014) is another widely used instrument which may be completed by parents, teachers, or clinicians to aid in the identification of ASD and quantify ASD related characteristics in six domains; 1: *Restricted/Repetitive Behaviours*, 2: *Social Interaction*, 3: *Social Communication*, 4: *Emotional Responses*, 5: *Cognitive Style* and 6: *Maladaptive Speech*. The instrument is designed to estimate global ASD severity and quantify atypicality in these subscale areas for individuals with ASD aged between three and 22 years. The GARS-3 is comprised of 58 items measuring to what extent a specified behaviour describes the child on a scale from 0 (*not at all like the individual*) to 3 (*very much like the individual*). Scaled scores are derived based on the sum of scores in each domain. In turn, these scores are used to generate an *Autism Index* which indicates the likelihood of ASD and support levels based on those outlined in the DSM-5 (American Psychiatric Association, 2013). At this clinic, the GARS-3 is completed by parents to supplement diagnostic assessment and feedback, and to guide intervention planning.

Intellectual Ability

Standardised IQ information derived using the Wechsler Intelligence Scale for Children (4th or 5th edition; Wechsler, 2003; 2014) or Wechsler Preschool and Primary Scale of Intelligence (3rd or 4th edition; Wechsler, 2002; 2012) was collected if available. As per the clinic's standard developmental assessment procedure, diagnosticians who are psychologists typically include an assessment of intellectual ability if the child is of sufficient age and able to engage, if time allows, and particularly if there are concerns about cognitive ability. An intellectual ability screen (i.e., the administration of a small number of subtests) is usually attempted. For the purpose of the present study, if a complete assessment necessary to obtain a Full-Scale Intelligence Quotient (FSIQ) was not conducted, prorated FSIQs were generated based on the subtest or composite scores reported. IQ information was available for

39.7% of children scored using the CARS2-ST, and 64.4% and 64.5% using the CARS2-HF and GARS-3, respectively. Comparisons of FSIQs were conducted based on the subsample for whom these data were available.

Statistical Analysis

A Bayesian parameter estimation approach was used for the analysis of these data. Hierarchical mixed linear modelling was selected over other Bayesian methods (e.g., analysis of covariance) so as to account for all sources of variability, including the overall severity of a child's ASD according to these instruments. The use of this model allowed the overall severity (model intercept) to vary by client (random effect).[16] Bayesian techniques more flexibly and naturally manage mixed-effects modeling than frequentist approaches (Gelman et al., 2014; Kruschke & Vanpaemel, 2015).

Many studies presenting ordinal data have used analyses which treat the variables as though they were continuous or nominal (Agresti, 2010).[17] In this study, a cumulative-normal threshold mixed-effects model (also known as an ordered probit model) was constructed for the major analyses (i.e., estimation of relative severity levels in each item) so as to best account for the ordinal nature of these data. This approach does not assume, for example, that a score of 3 on a CARS2 (ST or HF) item reflects atypicality that is twice as severe as that rated 1.5 in another child. Instead, this model estimates the mean and variance of a latent continuous severity variable conceptualised as underlying the ordinal responses. Thus, the sex/gender difference will be reflected in the difference in estimated mean of the latent variable and is best represented, as any difference between means, using a standardised mean

[16] The mathematical formulations underlying the models are presented in Appendix C.

[17] A regular mixed-effects Bayesian analysis was conducted first and compared with the results reported here. There were meaningful and obvious differences in the results between the two approaches, highlighting the importance of using analyses intended for the type of data collected (in this case, ordinal).

difference effect size (i.e., Cohen's *d*). Importantly, this type of analysis (for review, see Kruschke & Liddell, 2018) is more sensitive to interactions (the interaction between sex/gender and item is of primary interest in the item-by-item analysis) and avoids the issues associated with treating ordinal data as continuous and the relative lack of informativeness of purely categorical data analysis (e.g., chi-square or hierarchical loglinear analyses).

In order to estimate the probability that an average male or female would present with any atypicality in a specific area, the proportion of the estimated average latent distribution that lay above the estimated cut-off for a rating of 2 is presented. This model estimated the difference in probability that a male versus a female would score 2 or higher (a score of 2 indicates *Mild Abnormality* on the CARS forms, and that the behaviour is *Somewhat Like the Individual* on the GARS-3).[18]

Finally, I conducted *t*-tests on the CARS2 total scores (both ST and HF forms) and GARS index scores to compare the total scores of males and females. These analyses modelled the data, for males and females separately, as *t*-distributed. Using the estimated mean, scale, and normality[19] parameters from these analyses, I was able to calculate the proportion of the distribution (i.e., the estimated proportion of scores in a similar sample) that would fall above clinically relevant cut-offs. Thus, I estimated the proportion of males and females falling within each severity level according to the instruments.

---

[18] A score of 1 on the CARS2 forms (ST and HF) indicates *Age appropriate behaviour* or *No evidence of difficulty* and on the GARS-3, a score of 0 indicates that a behaviour is *Not at all like the individual* and 1, that a behaviour is *Not much like the individual*.

[19] The scale parameter reflects the standard deviation of scores (but the standard deviation for a *t*-distribution is undefined, thus the generic term scale) and the normality parameter reflects the extent to which the distributions shape differs from the normal distribution. As the parameter approaches infinity, the distribution approaches the normal distribution. (In NHST analyses where the *t*-distribution is used as a sampling distribution this parameter is known as the degrees of freedom.)

The aim was not to elucidate the causes of the difference or to identify whether or not it was a product of characteristics of the male versus female clients. Instead, my goal was to identify the items for which sex/gender differences arose on these common measures as used in clinical practice. Thus, I did not include as covariates any of the variables representing the clinical profile of the sample, such as age or cognitive ability.[20] Consequently, these results describe the sex/gender differences in profiles that clinicians encounter in the course of their clinical practice, rather than a controlled reflection of the true symptom severity of males versus females.

Results

The current investigation examined differences between autistic males and females in the severity of ASD characteristics in specific subdomains and behaviours using CARS2 (ST or HF) and GARS-3 profiles constructed during their diagnostic assessment. In doing so, two sets of information were gleaned: (a) the relative severity of ASD characteristics of males and females in specific areas, and (b) the nature of ASD traits that may be characteristic of a particular sex/gender.

Results are displayed in Figures 3.2-3.7 and Tables 3.1-3.7 which indicate the 80% highest density interval ($HDI_{80\%}$) of predicted values of the latent variable. The $HDI_{80\%}$ shows the range of values that are most credible for a particular parameter and span 80% of the distribution of credible values. Therefore, one can be 80% confident that the true value lies within this range. The region of practical equivalence (ROPE) was defined as ± 0.1 on the Cohen's *d* scale (Kruschke, 2018). The sign of the effect size indicates the direction of difference (positive for males, negative for females). That is, male sex/gender is assigned the

[20] For comparison, analyses including age and IQ as covariates were conducted in addition to those reported here. The differences between these sets of results were minimal and, in most cases, negligible.

positive direction (i.e., M > F = positive), and female sex/gender is assigned the negative direction (i.e., F > M = negative). A difference was judged to be meaningful if the $HDI_{80\%}$ lay entirely outside of the ROPE. This binary decision was supplemented with a quantification of the probability that the true difference in latent means was meaningful (i.e., the probability that the posterior distribution lay outside the ROPE) and denoted by $P_{(meaningful)}$ estimates.

Age and Intellectual Ability

Differences in age and intellectual ability between males and females were considered for children in each sample. Bayesian *t*-tests revealed no meaningful differences on these two variables between males and females that were likely to be of clinical significance (refer to Table 3.1). For children whose presentations were scored using the CARS2-HF form, there was probable evidence of a difference in age (older mean age in females; *M* difference = 0.63 years, $HDI_{80\%}$ = [0.21, 1.09]). However, there was meaningful evidence of a small difference in IQ (higher in females; *M* difference = 4.90, $HDI_{80\%}$ = [0.83, 8.91]). Although statistically meaningful, this difference was within the standard error of measurement (± 6; Wechsler, 2014) and is therefore not likely to be *clinically* meaningful. There was no strong evidence of a difference in IQ for children assessed using the CARS2-ST form. However, a small difference in age cannot be excluded with 80% confidence (higher in females; *M* difference = 0.41, $HDI_{80\%}$ = [0.03, 0.79]). For children whose presentations were scored against the GARS-3, there was equivocal evidence of a difference in both age and IQ.

Table 3.1

Sex/Gender Differences in Age and IQ of Children Assessed with Each Instrument

Instrument	Variable		Males	Females	*d* *M* [$HDI_{80\%}$]	$P_{(meaningful)}$
CARS2-ST		*n*	179	68		
	Age (yrs)	*M (SD)*	5.00 (2.31)	5.70 (2.71)	-0.26 [-0.49, -0.02]	-.81
		Range	2.00-15.75	2.16-12.83		
	IQ	*M (SD)*	90.04 (17.28)	90.33 (17.49)	-0.02 [-0.32, 0.27]	-.37
		Range	47-125	56-122		
CARS2-HF		*n*	214	109		
	Age (yrs)	*M (SD)*	9.27 (2.60)	9.92 (2.98)	-0.23 [-0.39, -0.07]	-.86
		Range	6.00-17.08	6.08-17.83		
	IQ	*M (SD)*	99.26 (12.45)	104.07 (13.86)	**-0.38 [-0.58, -0.18]**	**-.96**
		Range	80-136	80-133		
GARS-3		*n*	163	96		
	Age (yrs)	*M (SD)*	8.45 (3.53)	8.96 (3.34)	-0.15 [-0.32, 0.02]	-.65
		Range	2.16-17.25	2.83-16.33		
	IQ	*M (SD)*	95.03 (16.48)	97.63 (14.62)	-0.14 [-0.36, 0.07]	-.60
		Range	40-134	61-131		

*Note. N*s reflect total numbers of children scored according to each instrument, not only those for whom IQ data was available (this descriptive data is presented in Appendix B, Table B.2). Fifty-two children were scored against both a CARS2 form (ST or HF) and the GARS-3. Within each of the groups of children scored against the CARS2-ST and GARS-3, eight males and three females were identified as having full-scale IQs ≤ 70 (the approximate threshold for an intellectual disability). *d* reflects the effect size of the male-female mean difference. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and indicates higher age/IQ in females (depicted by the negative sign).

Item-Level Sex/Gender Differences

CARS2 Standard Form

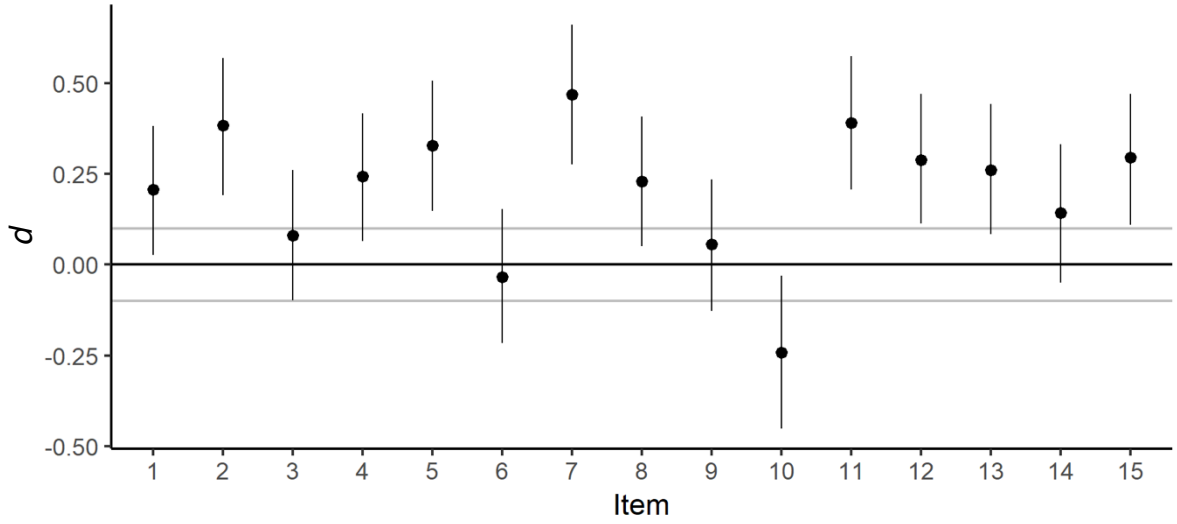
To explore the sex/gender differences in CARS2-ST item scores (i.e., ASD subdomains), hierarchical ordered probit models were fit to the ratings using separate models for the standard and high-functioning forms. These models estimate the latent distribution of severity underlying the ordinal ratings. Thus, the sex/gender difference is reflected in the difference in estimated means of the latent distributions for males versus females.[21]

I used mixed linear modeling and estimated scores for the items by considering the child's sex/gender, the item and the item by sex/gender interaction, allowing overall ASD severity to vary by participant. The most credible values for effect sizes of the difference in the latent means between males and females for each CARS2-ST item are presented in Figure 3.2. The strength of evidence of a sex/gender difference can be deduced from the position of the $HDI_{80\%}$ (error bars) relative to the ROPE (i.e., the region in which a sex/gender difference is too small to be meaningful, indicated by horizontal lines). Table 3.2 presents CARS2-ST item descriptions, effect sizes, and probabilities that the posterior distribution of each effect size lay outside the ROPE (indicating meaningful and negligible differences, respectively). For six CARS2-ST items, the $HDI_{80\%}$ of the effect size lay completely above the ROPE, meaning we can conclude with 80% confidence that there was a meaningful difference in atypicality between males and females. As the HDIs lay *above* the ROPE, atypicality was greater amongst males. From largest to smallest effect size, these items were: 7: *Visual response,* 2: *Imitation,* 11: *Verbal communication,* 5: *Object use,* 12: *Nonverbal communication,* and 15: *General impressions.* Considering that the CARS2-ST item scale

[21] For scores on each CARS2 form, I ran models *excluding* item 15: *General Impressions,* for the purpose of comparison with those presented here. Scores on this item did not meaningfully impact results. Thus, there is no evidence that sex/gender differences in item scores were driven by clinicians' overall impressions.

Figure 3.2

Plot of Effect Size of Sex/Gender Difference (Male - Female) in Estimated Mean of Latent Severity Distribution for Each CARS2-ST Item



Note. Error bars represent the $HDI_{80\%}$. Horizontal grey lines depict the ROPE's upper and lower bounds. Differences were judged to be meaningful when the $HDI_{80\%}$ lay entirely outside of the ROPE. Probable evidence of a difference is obtained when the $HDI_{80\%}$ excludes zero, but still overlaps the ROPE.

ranges from 1 (*no atypicality*) to 4 (*marked atypicality*), to determine the extent to which the results are clinically meaningful, it is valuable to consider them on the measurement scale (i.e., from largest to smallest of items with meaningful sex/gender differences; *M* difference = 0.21, $HDI_{80\%}$ = [0.12, 0.30] to 0.14, $HDI_{80\%}$ = [0.05, 0.23]). Therefore, differences of the magnitude found here were likely to be clinically meaningful. For each of these items, the probability that the true effect size of the difference between males and females was meaningful according to the ROPE criterion and reflected greater atypicality amongst males was ≥ 91.4% (i.e., probability that the true value lay above the ROPE). Therefore, we can have greater than 91.4% confidence that males had meaningfully greater atypicality in these areas (i.e., Items 2, 5, 7, 11, 12 and 15). For another four items, there was probable evidence of a sex/gender difference with greater atypicality amongst males: 1: *Relating to people*, 4:

Table 3.2

Descriptions and Estimates of Effect Sizes in Sex/Gender Differences in CARS2-ST Item Scores

CARS2-ST Item and Description (abbreviated)		*d* [$HDI_{80\%}$]	$P_{(meaningful)}$
1. Relating to people	Response to communication, initiation of interaction, aloofness/awareness of others	0.21 [0.03, 0.38]	.78
2. Imitation	Reliability, spontaneity, and immediacy of imitation, ability to imitate others	**0.38 [0.19, 0.57]**	**.98**
3. Emotional response	Appropriateness of emotion to situation, type/degree of emotional response	0.08 [-0.10, 0.26]	.44
4. Body use	Motor peculiarities and movement stereotypies, clumsiness, coordination	0.24 [0.07, 0.42]	.85
5. Object use	Degree of interest in objects, focus on parts of objects, repetitive or inappropriate use	**0.33 [0.15, 0.51]**	**.95**
6. Adaptation to change	Response to changes in routine, transitioning	-0.03 [-0.22, 0.15]	-.33
7. Visual response	Abnormality in eye contact, visual stereotypies, visual sensory behaviour	**0.47 [0.28, 0.66]**	**1.00**
8. Listening response	Auditory hyper-/hypo-sensitivity	0.23 [0.05, 0.41]	.83
9. Taste, smell, touch response and use	Response to sensory stimulation, use of these sensory modalities	0.06 [-0.13, 0.23]	.38
10. Fear or nervousness	Degree and context of anxious/nervous response	-0.24 [-0.45, -0.03]	-.81
11. Verbal communication	Unusual speech mannerisms, preoccupation with certain topics, repetitive speech	**0.39 [0.21, 0.57]**	**.98**
12. Nonverbal communication	Impairment in expression, interpretation of nonverbal communication	**0.29 [0.11, 0.47]**	**.91**
13. Activity level	Hyper/hypo-activity	0.26 [0.08, 0.44]	.88
14. Level and consistency of intellectual response	General level of intellectual functioning, consistency in cognitive abilities	0.14 [-0.05, 0.33]	.61
15. General impressions	Clinical impression of ASD severity	**0.30 [0.11, 0.47]**	**.92**

Note. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. Negative values indicate greater atypicality amongst females. *d* reflects the effect size of the male - female difference in the estimated latent means. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater atypicality in males; negative probabilities indicate greater atypicality in females).

Body use, 8: *Listening response*, and 13: *Activity level*. Specifically, an effect of exactly zero could be excluded with 80% confidence as in each case the $HDI_{80\%}$ did not cross zero. However, the possibility of a negligibly sized effect could not be ruled out with 80% confidence as, in each case the $HDI_{80\%}$ overlapped the ROPE. For each of these items, the probability that the true difference reflected meaningfully greater atypicality amongst males in these areas was between 78.2% and 88.0% (i.e., probability that the true value lay above the ROPE). Probable evidence of greater atypicality amongst females was only found for Item 10: *Fear or Nervousness.* This effect was moderate in strength, with a mean difference of -0.11 ($HDI_{80\%}$ = [-0.22, 0.02]), 80.6% probability of greater atypicality amongst females and 17.7% probability of no meaningful difference.

No meaningful sex/gender difference was found for the remaining four items, but the certainty of these estimates was insufficient to conclude a null effect with 80% confidence as the HDIs were contained entirely within the ROPE. These were Items 3: *Emotional response,* 6: *Adaptation to change,* 9: *Taste, smell and touch response and use,* and 14: *Level and consistency of intellectual response*. Only in Item 14 was the most probable effect a difference favouring males (rather than a negligibly-sized difference), although this evidence was equivocal.

Sex/Gender Differences in Probability of Presenting with Atypicality. In order to examine differences in the probability that the average male or female would present with any notable atypicality in a specific area, I calculated the proportion of the average estimated

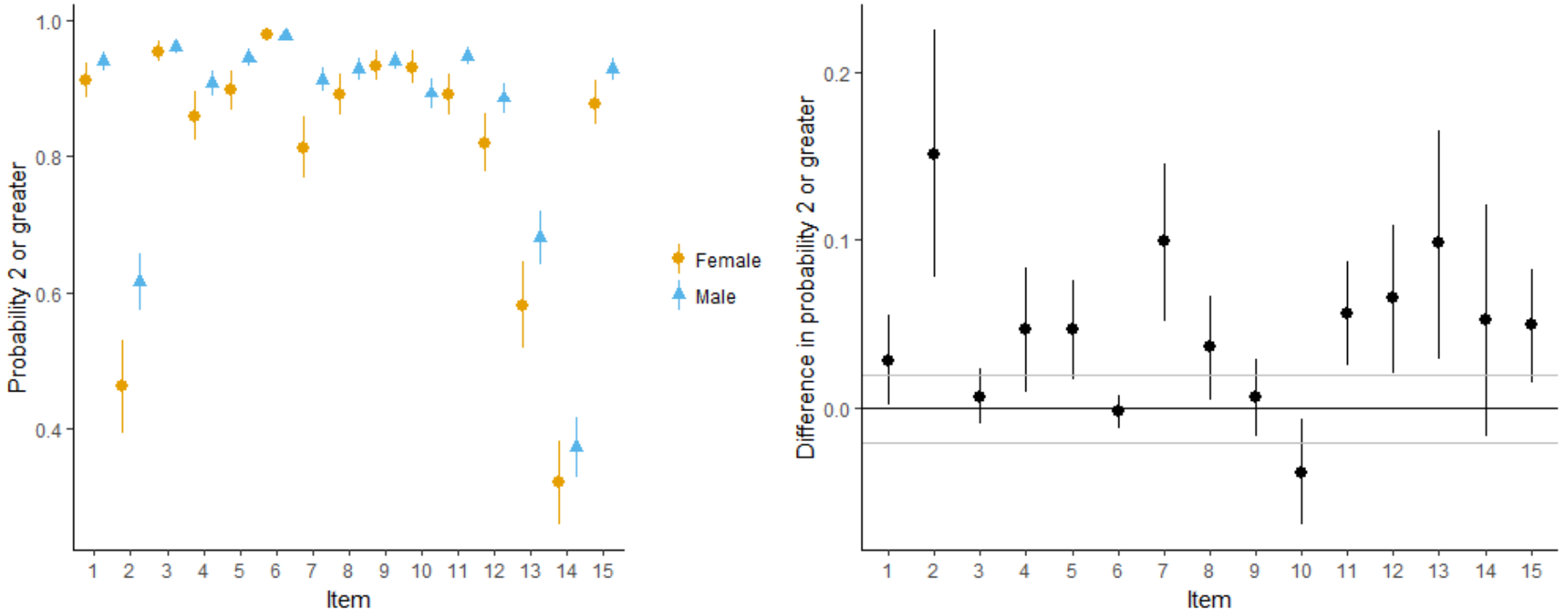
latent distribution that lay at or above the estimated cut-off for a rating of 2 (*Mild Abnormality*) for each item. The findings of this analysis are presented in Figure 3.3, which shows the probability of the average male or female being rated ≥ 2 on each item and the difference between these probabilities.

For three items (2, 7, and 13), evidence suggested that the average male was over 9.7% more likely than the average female to be rated as presenting with atypicality. Indeed, for item 2: *Imitation*, males were 15.1% more likely than females to obtain a score ≥ 2. Males were also between 5.0% and 7.0% more likely than females to present with atypicality in items 11, 12, and 14. Broadly, there was evidence that for eight items, we can have greater than 80% confidence that males were more than 2% more likely than females to present with atypicality (indeed, the minimum $P_{(meaningful)}$ was .84). In contrast, there was probable evidence of a difference favouring females in item 10: *Fear or nervousness,* in which females were slightly (3.8%) more likely than males to present with any atypicality.[22]

[22] Item-level estimated mean sex/gender differences in the latent variable and in the probability of a score ≥ 2 are presented in Appendix D (Table D.1).

Figure 3.3

Plots of the Probability of a Score ≥ 2 on each CARS2-ST Item by Sex/Gender (Left Panel) and of Sex/Gender Difference (Male - Female) in Probability of Obtaining ≥ 2 (Right Panel)

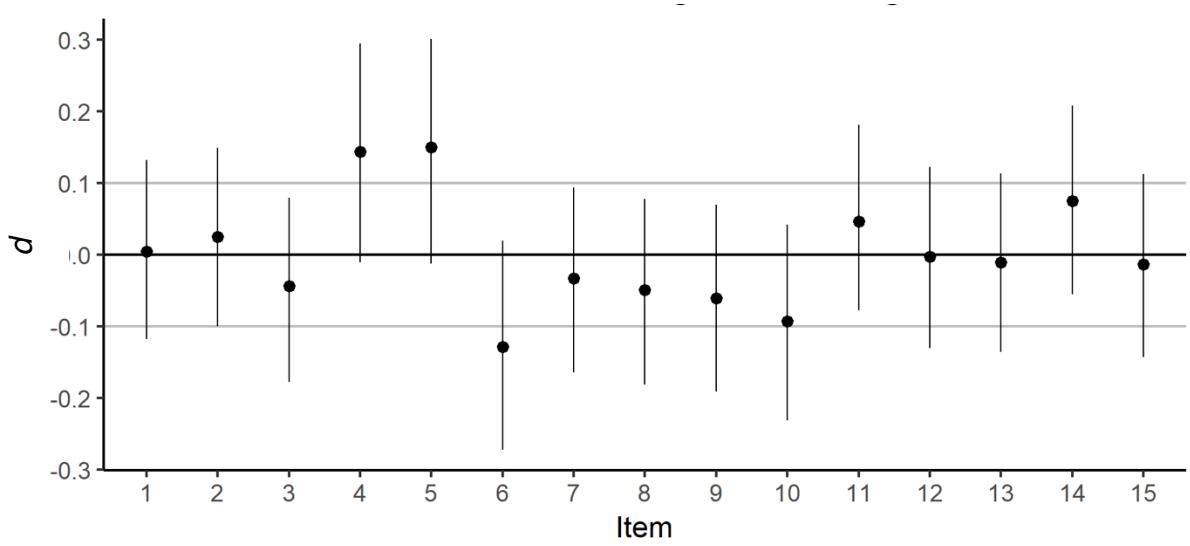


CARS2 High-Functioning Form

There was no evidence supporting a meaningful sex/gender difference for any single CARS2-HF item (refer to Figure 3.4 for a graphical summary of item-level differences and Table 3.3 for item descriptions and estimates of evidence strength). For 12 items, the most plausible estimated latent mean difference was within the ROPE and the balance of evidence supported no meaningful difference, albeit not with 80% confidence. For only three items was the best estimate outside the negligible range: 4: *Body use*, 5: *Object use in play,* and 6: *Adaptation to change/restricted interests*. Of these, only item 6 suggested greater atypicality amongst females. In all three of these cases, the $HDI_{80\%}$ did not exclude zero as a plausible estimate of the difference. Thus, although suggesting that there may be a sex/gender difference in ratings (i.e., $P_{(meaningful)}$ = 61.3 to 65.6% probability), these data were equivocal.

Figure 3.4

Plot of Effect Size of Sex/Gender Difference (Male - Female) in Estimated Mean of Latent Severity Distribution for each CARS2-HF Item



Note. Error bars represent the $HDI_{80\%}$ and horizontal grey lines depict the ROPE's upper and lower bounds. Differences were judged to be meaningful when the $HDI_{80\%}$ lay entirely outside of the ROPE. Weaker evidence of a difference is obtained when the $HDI_{80\%}$ excludes zero, but still overlaps the ROPE.

Table 3.3

Descriptions and Estimates of Effect Sizes in Sex/Gender Differences in CARS2-HF Item Scores

CARS2-HF Item and Description (abbreviated)		*d* [$HDI_{80\%}$]	$P_{(meaningful)}$	$P_{(within)}$
1. Social-emotional understanding	Understanding of non-verbal cues, perspectives of others	0.01 [-0.12, 0.13]	.16	.70
2. Emotional expression/ regulation	Appropriateness of type, degree of emotion, emotional regulation, understanding	0.03 [-0.10, 0.15]	.22	.69
3. Relating to people	Initiation of interaction, reciprocity of interactions	-0.04 [-0.18, 0.08]	-.30	.64
4. Body use	Motor peculiarities, movement stereotypies, clumsiness, fine/gross motor skills	0.14 [-0.01, 0.30]	.64	.38
5. Object use in play	Interest in toys or objects, repetitive/inappropriate use, imagination/spontaneity in play	0.15 [-0.01, 0.30]	.66	.34
6. Adaptation to change/restricted interest	Special and limited interests, rituals, routines, and ability to cope with change and transitions	-0.13 [-0.27, 0.02]	-.60	.40
7. Visual response	Abnormality in eye contact, gaze switching, visual stereotypies, visual sensory behaviour	-0.03 [-0.16, 0.09]	-.26	.66
8. Listening response	Auditory hyper-/hypo-sensitivity, response to name	-0.05 [-0.18, 0.08]	-.31	.63
9. Taste, smell, and touch response and use	Response to sensory stimulation, use of these sensory modalities	-0.06 [-0.19, 0.07]	-.35	.60
10. Fear or anxiety	Extent of unusual fear or anxiety relative to context	-0.09 [-0.23, 0.04]	-.47	.51
11. Verbal communication	Verbal oddities, conversation reciprocity	0.05 [-0.08, 0.18]	.30	.64
12. Non-verbal communication	Use of facial expression and gestures, response to non-verbal behaviour, joint attention	-0.00 [-0.13, 0.12]	-.16	.70
13. Thinking and cognitive integration	Attention to detail, weak central coherence	-0.01 [-0.14, 0.11]	-.18	.70
14. Level and consistency of intellectual response	Overall intellectual functioning, consistency in cognitive abilities	0.08 [-0.06, 0.21]	.41	.56
15. General impressions	Clinical impression of ASD severity	-0.01 [-0.14, 0.11]	-.19	.69

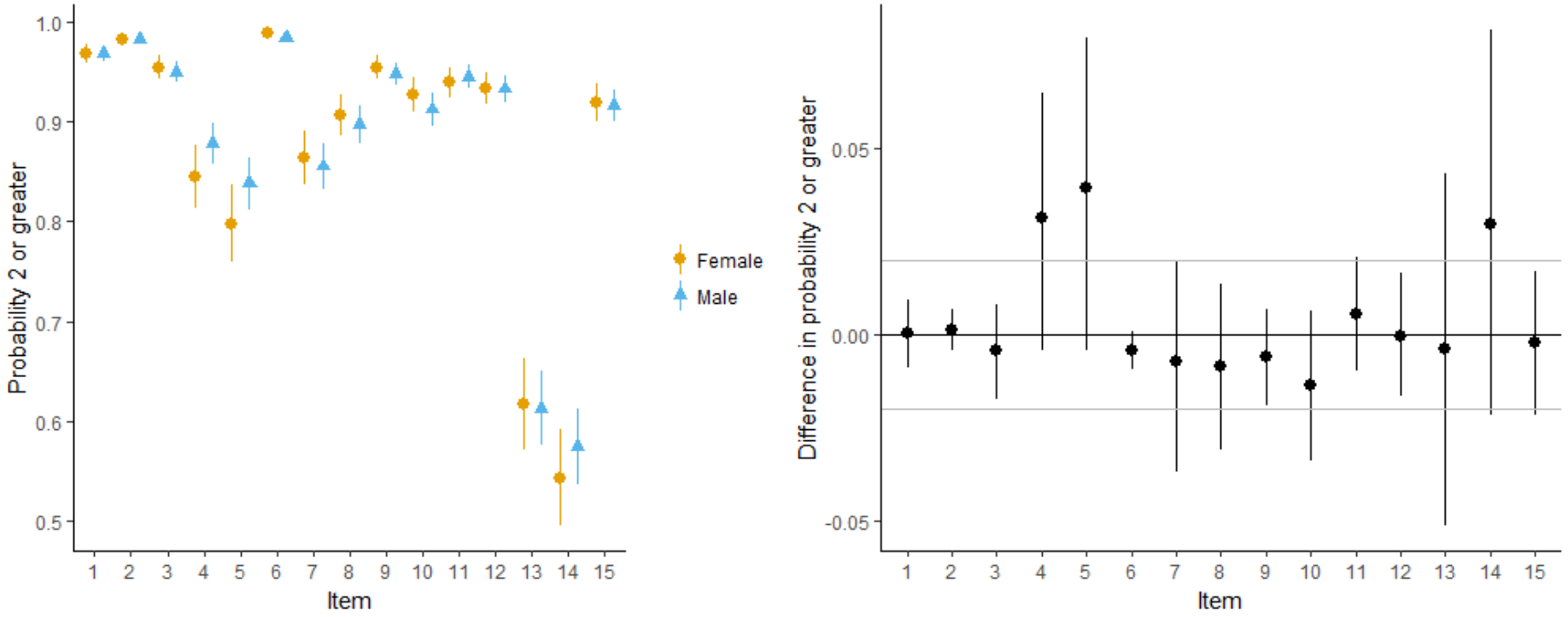
Note. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. Negative values indicate greater atypicality amongst females. *d* reflects the effect size of the male - female difference in the estimated latent means. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater atypicality in males; negative probabilities indicate greater atypicality in females). $P_{(within)}$ = probability that the true difference fell within the negligible range.

Sex/Gender Differences in Probability of Presenting with Atypicality. As shown in Figure 3.5, we can conclude that there was no sex/gender difference in the probability of a child receiving a score of 2 (*Mild Abnormality*) or higher in six of the items, and that the balance of evidence favours no meaningful difference for five others. For the remaining items, evidence of a sex/gender difference was equivocal.[23]

[23] Item-level estimated mean sex/gender differences in the latent variable and in the probability of a score ≥ 2 are presented in Appendix D (Table D.2).

Figure 3.5

Plots of the Probability of a Score ≥ 2 on each CARS2-HF Item by Sex/Gender (Left Panel) and of Sex/Gender Difference (Male - Female) in Probability of Obtaining ≥ 2 (Right Panel)



GARS-3

I also fit hierarchical ordered probit models to children's scores on the GARS-3 items, each reflecting a particular ASD behaviour. Refer to Figure 3.6 for a graphical summary of item-level differences and Table 3.4 for items in which a difference of zero could be excluded with 80% confidence.[24]

Meaningful sex/gender differences were found in the scores of six specific GARS-3 items. In addition, probable differences (i.e., in which the $HDI_{80\%}$ overlapped the ROPE but not zero) were found for a further 23 items. For all items in which a meaningful or probable sex/gender difference was found, males were rated as having more severe atypicality.

The strongest evidence for more severe atypicality amongst males was found in Items 48: *Displays superior knowledge or skill in specific subjects* ($P_{(meaningful)}$ = 98.8%) and 50: *Shows an intense, obsessive interest in specific intellectual subjects* ($P_{(meaningful)}$ = 97.9%), both in the domain of *Cognitive Style.* Meaningful differences were also found for Items 7: *Makes high-pitched sounds/other vocalisations for self-stimulation*, 8: *Uses toys or objects inappropriately*, 17: *Does not follow others' gestures (cues) to look at something*, and 26: *Fails to engage in creative, imaginative play*. In these areas, the probability of greater atypicality amongst males ranged from 85.1% to 95.0%. When considered on the GARS-3 rating scale (0-3) each of these differences was moderate: *M* difference between 0.25 ($HDI_{80\%}$ = [0.11, 0.39]) for Item 25, and 0.42 ($HDI_{80\%}$ = [0.21 to 0.61]) for Item 48.

For 23 items, there was evidence of a probable sex/gender difference with greater severity in males ($P_{(meaningful)}$ = 64.4% to 85.0%), although in these cases a negligible effect could not be excluded with 80% confidence. The balance of probabilities suggested a

[24] A summary of all items and estimates of sex/gender differences is presented in Appendix D (Table D.3).

Figure 3.6

Plot of Effect Size of Sex/Gender Difference (Male - Female) in Estimated Mean of Latent Severity Distribution for Each GARS-3 Item

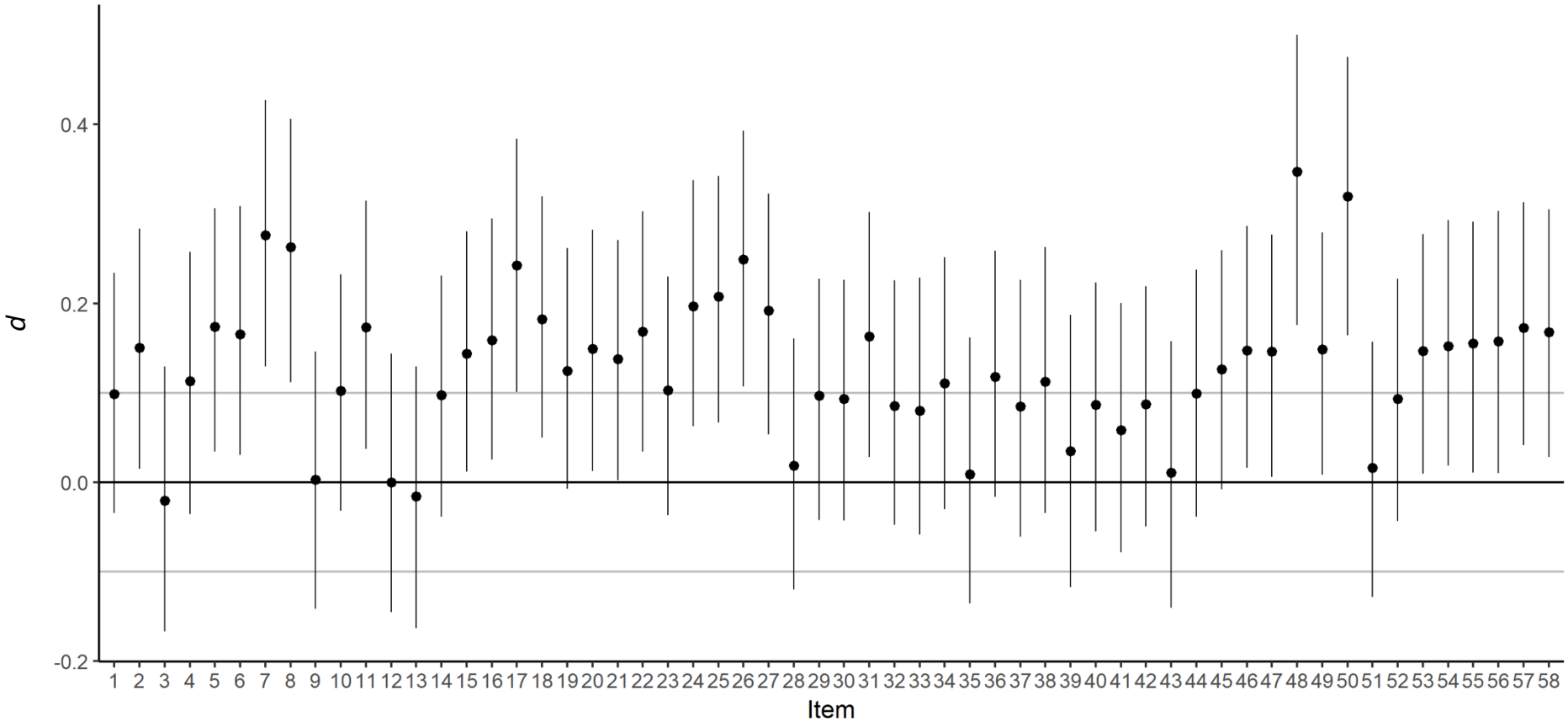


Table 3.4

Descriptions and Estimates of Effect Sizes in Sex/Gender Differences for GARS-3 Item Scores in Which a Difference of Zero Was Excluded with 80% Confidence

GARS-3 Items (abbreviated)	*d* [$HDI_{80\%}$]	$P_{(meaningful)}$
Restricted/Repetitive Behaviours (RB)		
2. Is preoccupied with specific stimuli that are abnormal in intensity	0.15 [0.02, 0.28]	.69
5. Makes rapid lunging, darting movements when moving from place to place	0.17 [0.04, 0.31]	.76
6. Flaps hands or fingers in front of face or at sides	0.17 [0.03, 0.31]	.73
7. Makes high-pitched sounds or other vocalisations for self-stimulation	**0.28 [0.13, 0.43]**	**.95**
8. Uses toys or objects inappropriately (e.g., spins cars, takes action toys apart)	**0.26 [0.11, 0.41]**	**.94**
11. Repeats unintelligible sounds (babbles)	0.17 [0.04, 0.32]	.76
Social Interaction (SI)		
15. Pays little or no attention to what peers are doing	0.14 [0.01, 0.28]	.67
16. Fails to imitate other people in games or learning	0.16 [0.03, 0.30]	.72
17. Does not follow others' gestures to look at something	**0.24 [0.10, 0.38]**	**.92**
18. Seems indifferent to other person's attention	0.18 [0.05, 0.32]	.79
20. Displays little or no excitement in showing toys or objects to others	0.15 [0.01, 0.28]	.68
21. Seems uninterested in pointing out things in the environment to others	0.14 [0.00, 0.27]	.64
22. Seems unwilling or reluctant to get others to interact with him/her	0.17 [0.03, 0.30]	.75
24. Displays little/no reciprocal communication	0.20 [0.06, 0.34]	.83
25. Doesn't try to make friends with people	0.21 [0.07, 0.34]	.85
26. Fails to engage in creative, imaginative play	**0.25 [0.11, 0.39]**	**.92**
27. Shows little or no interest in other people	0.19 [0.05, 0.32]	.82
Social Communication (SC)		
31. Has difficulty identifying when someone is teasing	0.16 [0.03, 0.30]	.73
Cognitive Style (CS)		
46. Attaches very concrete meaning to words	0.15 [0.02, 0.29]	.68

GARS-3 Items (abbreviated)	*d* [$HDI_{80\%}$]	$P_{(meaningful)}$
47. Talks about a single subject excessively	0.15 [0.01, 0.28]	.67
48. Displays superior knowledge or skill in specific subjects	**0.35 [0.18, 0.50]**	**.99**
49. Displays excellent memory	0.15 [0.01, 0.28]	.68
50. Shows an intense, obsessive interest in specific intellectual subjects	**0.32 [0.16, 0.48]**	**.98**
Maladaptive Speech (MS)		
53. Repeats words out of context (heard earlier)	0.15 [0.01, 0.28]	.67
54. Speaks (or signs) with flat tone, affect	0.15 [0.02, 0.29]	.70
55. Uses 'yes' or 'no' inappropriately	0.16 [0.01, 0.29]	.70
56. Uses 'he'/ 'she' instead of 'I' when referring to self	0.16 [0.01, 0.30]	.70
57. Speech is abnormal in tone, volume or rate	0.17 [0.04, 0.31]	.77
58. Utters idiosyncratic words or phrases that have no meaning to others	0.17 [0.03, 0.31]	.74

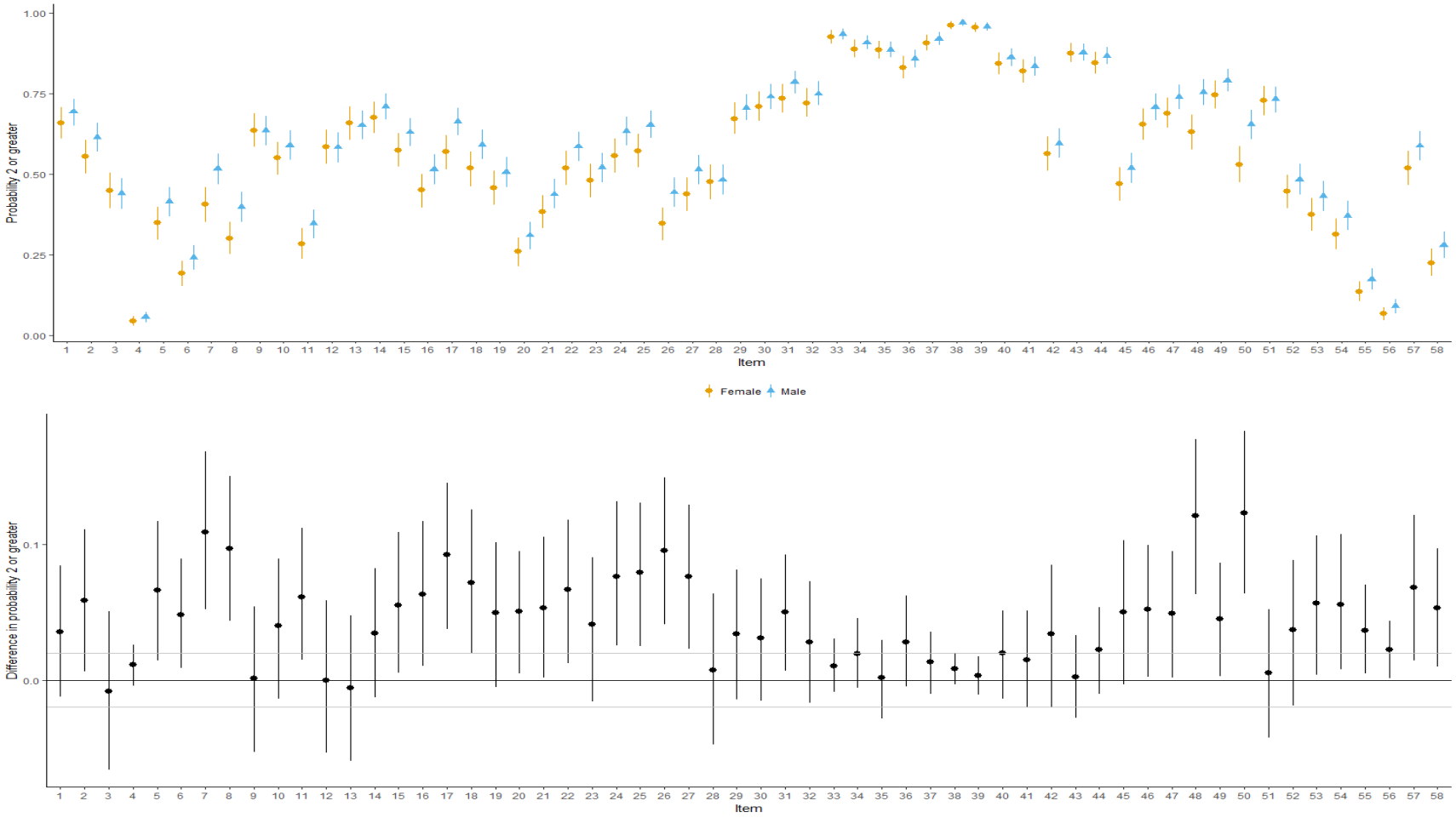
Note. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. Positive values indicate greater atypicality amongst males. *d* reflects the effect size of the male-female difference in the estimated latent means. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater atypicality in males).

negligibly sized difference for a further eight items ($P_{(within\ ROPE)}$ = 57.3% to 62.6%): Item 3: *Stares at objects for at least 5 seconds,* 12: *Shows unusual interest in sensory aspects of objects*, 13: *Displays ritualistic or compulsive behaviours*, 25: *Doesn't try to make friends with people*, 28: *Responds inappropriately to humorous stimuli*, 35: *Doesn't understand that people have thoughts and feelings different from his or hers*, 39: *Temper tantrums when frustrated*, 43: *Temper tantrums when doesn't get his or her way,* and 51: *Makes naïve remarks (unaware of reaction produced in others)*. This suggests atypicality levels were consistent between males and females in the sample in these specific behaviours.

Sex/Gender Differences in Probability of Presenting with Atypicality. As in the CARS2 forms, I considered the difference in probability that the average male or female would present with any atypicality in each item on the GARS-3 (refer to Figure 3.7). There

Figure 3.7

Plots of the Probability of a Score ≥ 2 on each GARS-3 Item by Sex/Gender (Upper Panel) and of Sex/Gender Difference (Male - Female) in Probability of Obtaining ≥ 2 (Lower Panel)



were nine items for which the probability of a rating of 2 (*Somewhat Like the Individual*) or higher was meaningfully greater for the male than female average: Items 7, 8, 17, 24-27, 48, and 50 (92.0-99.7% probability that the true difference favoured males and was greater than 2%). In Items 7, 48, and 50, the evidence suggested that the average male was 10.9-12.3% more likely to present with atypicality than the average female. The sex/gender difference in probabilities was slightly lower for Items 8, 17, and 26, in which the average male was 9.0-10.0% more likely than the average female to be rated a score of 2 or higher. Broadly, there was evidence that for 26 items, we can have greater than 79.9% confidence that the average male was at least 2% more likely to be rated as displaying atypicality than the average female. There was no evidence that the average female was more likely than the average male to be rated a score of 2 or higher on any item.

Sex/Gender Differences in Overall Severity Levels

CARS2-ST

Although not a primary focus of this investigation, total scores (i.e., the summation of all item scores) were compared for males and females in order to quantify any difference in overall ASD severity level. Bayesian *t*-tests found robust evidence of a difference in overall CARS2-ST scores between males (M = 33.36, $HDI_{80\%}$ = [33.00, 33.68], SD = 2.93, $HDI_{80\%}$ = [2.65, 3.19]) and females (M = 31.94, $HDI_{80\%}$ = [31.55, 32.31], SD = 2.15, $HDI_{80\%}$ = [1.77, 2.46]). The difference in means was 1.41 ($HDI_{80\%}$ = [0.91, 1.91]), with a large effect size of d = 0.79 ($HDI_{80\%}$ = [0.42, 1.20]; $P_{(meaningful)}$ = 98.8%).

I also used the distributions estimated in the above analysis to predict the probability that a male or female would reach a given severity threshold for overall ASD severity level using this measure (Table 3.5).[25] A greater proportion of males (M = 0.85, $HDI_{80\%}$ = [0.83,

[25] As there are two sets of severity thresholds on the CARS2-ST (i.e., children above or below 13 years), scores were stratified according to the child's age. There were insufficient older children to

0.88]) than females (M = 0.80 [0.75, 0.85]) were predicted to score at or above the threshold for *Mild-moderate* ASD (30 or higher on the ST form). Evidence indicated a 75.9% probability that this difference was meaningful (difference in Ms = 0.05, $HDI_{80\%}$ = [0.00, 0.11]). A greater proportion of males (M = 0.13, $HDI_{80\%}$ = [0.10, 0.16]) than females (M = 0.03 [0.01, 0.04]) were also predicted to score within the *Severe* range (≥ 37; difference in Ms = 0.10, $HDI_{80\%}$ = [0.07, 0.13]), with unequivocal evidence that this difference was meaningful ($P_{(meaningful)}$ = 99.9%).

CARS2-HF

Bayesian *t*-tests found no meaningful evidence of a difference in the total CARS2-HF scores between males (M = 32.94, $HDI_{80\%}$ = [32.69, 33.20], SD = 2.66, $HDI_{80\%}$ = [2.44, 2.87]) and females (M = 32.72, $HDI_{80\%}$ = [32.34, 33.11], SD = 2.90, $HDI_{80\%}$ = [2.59, 3.23]). The estimated difference in means was 0.22, $HDI_{80\%}$ = [0.23, 0.66], with an effect size of d = 0.08 ($HDI_{80\%}$ = [-0.08, 0.24]; $P_{(within\ ROPE)}$ = 81.6%). Thus, the balance of probabilities suggested there was no sex/gender difference in total scores on the HF form.

Similarly, analysis of overall scores revealed no meaningful difference between the proportions of males and females predicted to reach the CARS2-HF clinical severity thresholds (refer to Table 3.5). Specifically, males and females were equally likely to achieve a score of 28 or higher (HF threshold for *Mild-moderate* ASD; M difference = 0.02, $HDI_{80\%}$ = [0.00, 0.04]). This was also the case for a score of 34 (threshold for *Severe* ASD; M difference = 0.02, $HDI_{80\%}$ = [-0.05, 0.07]).

GARS-3

Bayesian *t*-tests revealed no evidence of a meaningful difference in overall GARS-3 index scores between males (M = 98.17, $HDI_{80\%}$ = [96.39, 99.68], SD = 16.11, $HDI_{80\%}$ =

---

conduct a meaningful analysis (n = 3) and therefore only scores of younger children (i.e., under 13 years old) are presented here.

Table 3.5

Estimates of Sex/Gender Differences in the Proportion of Children Meeting CARS2 (ST and HF) Severity Thresholds

Form and Severity Threshold		Proportion Reaching Threshold *M* [$HDI_{80\%}$]			
		Males	Females	Difference	$P_{(meaningful)}$
ST	Mild-Moderate (30)	0.85 [0.83, 0.88]	0.80 [0.75, 0.85]	0.05 [0.00, 0.11]	.76
	Severe (37)	0.13 [0.10, 0.16]	0.03 [0.01, 0.04]	**0.10 [0.07, 0.13]**	**1.00**
HF	Mild-Moderate (28)	0.96 [0.95, 0.97]	0.94 [0.92, 0.96]	0.02 [0.00, 0.04]	.50
	Severe (34)	0.35 [0.31, 0.38]	0.33 [0.28, 0.38]	0.02 [-0.05, 0.07]	.46

Note. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. Negative values indicate greater probability amongst females. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater atypicality in males; negative probabilities indicate greater atypicality in females).

[14.91, 17.30]) and females (*M* = 96.32, $HDI_{80\%}$ = [94.51, 98.11], *SD* = 13.56, $HDI_{80\%}$ = [12.31, 14.97], *M* difference = 2.45, $HDI_{80\%}$ = [0.62, 4.19]). However, we cannot conclude that the effect size was negligible with 80% confidence (*d* = 0.12 [-0.04, 0.28], $P_{(meaningful)}$ = 55.7%).[26]

[26] There are two major reasons why we might observe no evidence of a meaningful difference in index scores despite consistently higher estimated latent means among males on a large number of GARS-3 items. Firstly, the calculation of index scores involves individual item scores being aggregated and then categorised to produce a scaled score in each domain. This can reduce the extent to which small differences in item scores manifest in the scaled scores and therefore the index score. Second, the item-by-item analyses specifically model the latent variable underlying the ordinal ratings. One of the reasons that this is a superior approach to analysing the categorical ratings themselves is that it can detect differences in the latent variable that can be muted in the categorical

Table 3.6 presents the proportions of males and females predicted to reach each clinical threshold defined by the GARS-3. There was no meaningful difference in the predicted proportion of males and females scoring 55 to 70 (*Level 1: ASD Probable*; *M* difference = 0.00, $HDI_{80\%}$ = [-0.01, 0.00]). This was also the case for a score of 71 to 100 (*Level 2: ASD Very Likely*; *M* difference = -0.01, $HDI_{80\%}$ = [-0.04, 0.01]). There was weak evidence ($P_{(meaningful)}$ = 79.5%) of a meaningfully higher proportion of males scoring 101 or above (threshold for *Level 3: ASD Very Likely*; *M* difference = 0.06, $HDI_{80\%}$ = [0.00, 0.13]).

Table 3.6

Estimates of Sex/Gender Differences in the Proportion of Children Meeting Severity Thresholds (GARS-3)

Severity Level	Proportion Reaching Threshold *M* [$HDI_{80\%}$]			$P_{(meaningful)}$
	Males	Females	Difference	
Level 1: Probable ASD (55-70)	0.99 [0.99, 1.00]	0.99 [0.99, 1.00]	0.00 [-0.01, 0.00]	.00
Level 2: ASD Very Likely (71-100)	0.95 [0.94, 0.97]	0.96 [0.95, 0.98]	-0.01 [-0.04, 0.01]	-.37
Level 3: ASD Very Likely (≥101)	0.43 [0.39, 0.47]	0.37 [0.32, 0.42]	0.06 [0.00, 0.13]	.80

Note. Negative values indicate greater probability amongst females. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater atypicality in males; negative probabilities indicate greater atypicality in females).

ratings. Thus, a sex/gender difference could be present in the latent severity assessment for every item, but if it is small relative to the width of an ordinal category, that difference may be completely masked by the requirement to select one of a small number of ordinal response categories. In other words, even though a male-female pair may be assessed at slightly different levels of severity, in most cases the same ordinal response will best reflect the (different) underlying levels of severity. Thus, a consistent difference in the latent variable may not translate into a similarly strong difference in the categorical responses.

Domain-Level Sex/Gender Differences. The GARS-3 items are divided into six domains (refer to Table 3.7). Sex/gender differences in GARS-3 domain scores were analysed in order to estimate whether differences in severity of specific ASD characteristics (presented in the item-level analysis above) generalised to the domains in which these characteristics fall. Bayesian *t*-tests conducted on the scaled scores between males and females identified no meaningful sex/gender differences in scores in any of the six GARS-3 domains. However, there were probable differences in three domains (indicating greater atypicality amongst males): *Restricted/repetitive behaviours*, *Social interaction*, and *Maladaptive speech*. Although a difference of exactly zero could be excluded as a plausible difference in each case, a negligible difference could not be excluded with 80% confidence. For the remaining domains, the evidence of a sex/gender difference was equivocal and neither zero, nor a negligible effect in either direction could be eliminated with 80% confidence.

Table 3.7

Sex/Gender Differences in Severity Ratings in GARS-3 Domains

GARS-3 domain	*M* [$HDI_{80\%}$]		*d* [$HDI_{80\%}$]	$P_{(meaningful)}$
	Males	Females		
Restricted/repetitive behaviours	8.86 [8.61, 9.17]	8.39 [8.11, 8.72]	0.19 [0.03, 0.35]	.76
Social interaction	9.02 [8.75, 9.27]	8.36 [8.00, 8.72]	0.24 [0.08, 0.42]	.86
Social communication	8.54 [8.30, 8.81]	8.64 [8.34, 9.00]	-0.04 [-0.21, 0.13]	-.33
Emotional regulation	12.04 [11.70, 12.31]	11.90 [11.56, 12.22]	0.06 [-0.13, 0.26]	.40
Cognitive style	11.40 [11.17, 11.66]	11.03 [10.73, 11.32]	0.17 [0.00, 0.34]	.70
Maladaptive speech	9.49 [9.17, 9.78]	8.87 [8.54, 9.21]	0.22 [0.05, 0.39]	.82

Note. Negative values indicate greater probability amongst females. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater atypicality in males; negative probabilities indicate greater atypicality in females).

Discussion

The purpose of this study was to identify whether some of the commonly used ASD measurement instruments (i.e., CARS2-ST and HF; and GARS-3) were able to identify sex/gender differences in the presentation of ASD at the level of specific behaviours and groups of related behaviours (falling within particular subdomains) as a possible factor contributing to the underdiagnosis of ASD among females. Unlike previous studies, I also identified sex/gender differences in the probability of a child presenting with both typical (i.e., an absence of impairment) and atypical behaviours. Item-level scores were collected from profiles of children with ASD according to the CARS2 (ST, *n* = 247; and HF, *n* = 323) and GARS-3 (*n* = 259) compiled at the time of their ASD diagnosis. Meaningful sex/gender differences were found in six CARS2-ST items (subdomains) and six GARS-3 items

(specific behaviours), with some differences falling within the social communication domain, but the majority within the RRBI domain. Males scored higher than females in each of these items, and the average autistic female was more likely to present with developmentally typical behaviour in these same areas. Interestingly, females tended to demonstrate greater atypicality only in behaviours associated with fear/nervousness according to the CARS2-ST, but this was not reflected in the GARS-3 or CARS2-HF results. Indeed, no meaningful sex/gender differences were identified in any CARS2-HF items. Although not a particular focus of this study, overall ASD severity was similar for males and females scored using the CARS2-HF and GARS-3, but males were rated as having more difficulty on average and the average male was more likely to be rated as having *Severe* ASD on the CARS2-ST than the average female. Here, these findings are discussed in the context of previous literature and their implications for assessment.

The distribution of sex/gender differences in the severity of ASD characteristics found here is generally consistent with the theoretical framework and hypothesis of Lai et al. (2015). These authors suggested that sex/gender differences in ASD presentations may be most apparent within specific behaviours and subdomains in which specific behaviours are organised, rather than in the broad domains which define ASD (i.e., social communication difficulties and RRBIs). In the present study, there was evidence of meaningfully greater atypicality amongst males in six specific behaviours outlined on the GARS-3: superior knowledge of specific subjects, intense interests, failure of imaginative play, poor interpretation of nonverbal gestures, inappropriate use of toys, lack of attempt to make friends with others, and the production of noises for self-stimulation. There was weaker evidence of greater difficulty for males (compared to the females) for a further three behaviours pertaining to reciprocal conversation and social interest. In addition to identifying the above, this study extends previous literature by suggesting that, not only may females

with ASD show less atypicality than males in these areas, females may also be more likely than males to present with developmentally appropriate behaviour. Therefore, difficulties in the areas listed above may be less common among females with ASD than their male counterparts.

Sex/gender differences were also found in ASD characteristic severity in six CARS2-ST subdomains, most of which were consistent with the GARS-3 item-level results (i.e., Item 12: *Non-verbal behaviour*; Item 5: *Object use*; and Item 11: *Verbal communication;* or speech stereotypies/repetitiveness). However, greater atypicality in males in *Imitation* (Item 2) and *Visual response* behaviours (Item 7), which most differentiated males and females, were found only on the CARS2-ST form and not on the GARS-3. It is likely that differences in the specificity of items may contribute to differences observed across the instruments (i.e., more specific behaviours are assessed on the GARS-3 and behaviours are considered more broadly on each CARS2-ST item). Similarly, the clustering of particular behaviours on the CARS2-ST (e.g., abnormal eye contact, visual stereotypes and visual sensory behaviour, represented by Item 7: *Visual response*) may contribute to differences observed between the CARS2-ST and GARS-3 results. The specific behaviours in which males demonstrated more atypicality are largely consistent with previous literature examining sex/gender differences in specific ASD behaviours according to parent report (e.g., Antezana et al., 2018; Beggiato et al., 2017; Hiller et al., 2014). However, in contrast with these studies, no evidence of a meaningful sex/gender difference was found for hand and finger mannerisms or compulsive behaviours, perhaps due to differences in the operational definitions between instruments. In conjunction with sex/gender differences on specific items, weaker evidence suggested greater atypicality among males in three domains of the GARS-3: *Restricted/repetitive behaviours*, *Social interaction* and *Maladaptive speech*. Together, the results presented above are partially consistent with van Wijngaarden-Cremers and colleagues' meta-analysis (2014), in which

overall RRBI symptoms were found to be more severe among males. In particular, many of the specific behaviours (e.g., unusual vocalisations, use of objects, and restricted interests) and subdomains (e.g., visual behaviours) identified in the present study as being more pronounced among males can be classified under the RRBI umbrella. However, the lack of overall sex/gender difference in social communication, despite sex/gender differences identified in some studies, suggests that these differences may be dependent on the age or cognitive ability of participants, or, pertinent to this study, the specificity by which social difficulties are defined by the instruments used.

There are a number of discrepancies between the results of this study and those of Kumazaki et al. (2015), despite both using the CARS, albeit different versions.[27] For instance, I found a smaller, but still meaningful sex/gender difference in *Object use* (d = .33, compared to d = 1.11), and a smaller difference in *Body use*, (d = 0.24, narrowly below the meaningful criterion; compared to d = 0.68). The effect sizes were similar for *Imitation*, *Nonverbal communication*, and *Verbal communication*, but while differences emerged as meaningful in the present study, they were not 'statistically significant' in the Japanese study. Effects in opposite directions were observed for *Relating to people, General impressions,* and *Visual response,* and in contrast to the Japanese study, the balance of evidence in the present study suggested no difference in *Taste, smell and touch response and use.* The difference in results may be due to linguistic or operational differences between the forms used, statistical approach and/or sample size, or the difference in the average age of the children recruited

[27] To compare my findings with those of Kumazaki et al. (2015), I calculated an effect size (Cohen's *f*) from their reported ANCOVA results. As *f* is on the scale of half Cohen's *d,* I doubled the Cohen's *f* effect sizes for comparison with my results. Of course, this method produces only a point estimate without appropriately indicating the uncertainty in that estimate, so some caution must be exercised in making these comparisons. A summary of the comparisons between my findings and those of Kumazaki et al. (2015) are presented in Appendix D, Table D.4.

(2.17 years younger in the present study). It is also possible that cultural differences in normative gender roles, symptom expression, and interpretation (Matson et al., 2017) may account for the discrepancies.

For only one item, Item 10: *Fear/nervousness* (i.e., the degree and context of anxious/nervous response; CARS2-ST) was there evidence of greater atypicality amongst females in the present study. Although it is not possible to separate the specific behaviours that make up this item (e.g., difficulty regulating emotions and/or failure to demonstrate adaptive fear responses), these data suggest that atypicality in anxiety-related behaviours may constitute an element of the female ASD presentation, at least in how it is perceived by parents. While sex/gender differences in the relative severity of anxiety remain inconsistent in the broader literature, evidence has suggested that females may be less likely than males to externalise but more likely to internalise their distress (Hull et al., 2016; Mandy et al., 2011). Therefore, while no meaningful sex/gender differences were found in the GARS-3 items relating to emotional responses, this was thought to be the result of the insensitivity of the items to the *type* of atypical emotional responses (i.e., externalising or internalising) in their phrasing (e.g., *becomes upset* or *responds negatively*).

While autistic females may experience greater difficulty with anxiety than males, the current study suggests that a female advantage in imitation skills may be present from early childhood. In particular, substantially lower atypicality amongst females in Item 2: *Imitation* (CARS2-ST) and a 15.1% lower probability of presenting with any atypicality compared to males, supports previous literature (e.g., Backer van Ommeren et al., 2017). This female advantage in imitation may assist in explaining the greater camouflaging of ASD difficulties observed among females (Hull et al., 2018; Lai et al., 2016), and particularly the component of *assimilation* with typically developing peers. Hull et al. (2017) recently conceptualised assimilation as a combination of masking (concealment of ASD traits) and compensation

(strategies to aid in performing a "neurotypical role"), which together contribute to camouflaging. In addition to superior imitation abilities, social motivation, which was more likely to be considered developmentally typical among females compared to males, may also contribute to the motivation of some females to conceal ASD difficulties. Moreover, even if males and females were generally equally equipped regarding social and intellectual abilities, females may devote greater effort in the pursuit of camouflaging due to their desire to fit in with peers. In turn, this may contribute to greater camouflaging among autistic females compared to males (for a recent review, see Hull et al., 2020).

CARS2-High Functioning Form

There was no meaningful evidence of sex/gender differences on any CARS2-HF items; although for five items, the balance of evidence suggested a negligible difference (i.e., ~70% probability that the true difference fell within the negligible range). For the remaining items, evidence of a sex/gender difference was equivocal. The lack of meaningful sex/gender differences within RRBI subdomains was, on the surface, somewhat unexpected and appears to contrast with the results of van Wijngaarden-Cremers and colleagues' meta-analysis (2014), which showed greater RRBI difficulty among boys than girls (std. mean diff. = 0.19, $CI_{95\%}$ = [0.06, 0.32]). However, given their estimated effect size was small (indeed, it did not meet the typical minimum cut-off for a small effect of 0.20; Cohen, 1988), my observation of negligible to small size effects is similar. For example, while equivocal according to the criterion used in the present study, the effect sizes found for *Body use* and *Object use in play* (d = 0.14, $HDI_{80\%}$ = [-0.01, 0.30] and d = 0.15, $HDI_{80\%}$ = [-0.01, 0.30], respectively), were similar in magnitude to those reported in the meta-analysis. Together, these results suggest that, if there is a consistent non-negligibly sized difference, it is likely to be small. Having said this, small effects may still be clinically significant, either at the population level or for an individual who is at or narrowly below the threshold for ASD diagnosis.

The results of the present study lend some support to the hypothesis that females' social difficulties may emerge later in life and 'catch up' to males by late childhood, particularly when social demands outweigh social skills (Mandy et al., 2018). Children whose presentations were scored using the CARS2-HF form were approximately four years older than those scored using the CARS2-ST form. Sex/gender differences in items relating to social difficulties were more pronounced on the ST form (i.e., among younger children). While it is possible that the discrepancy in social skills difficulties may reduce as children with ASD approach adolescence (i.e., autistic females' difficulties grow more similar to those of males), it seems unlikely that this would explain the contrasting findings presented here given that younger females (scored on the CARS2-ST) demonstrated enough social difficulty to receive an ASD diagnosis.

Higher cognitive ability may be protective of ASD in females and/or allow them to better manage their condition, contributing to underdiagnosis and therefore the markedly higher prevalence ratio of 'higher functioning' males to females (Hull & Mandy, 2017). If higher cognitive ability is necessary for camouflaging, then the lack of meaningful sex/gender differences on the CARS2-HF could reflect the issue of considering only a diagnosed sample as, by definition, diagnosed females show sufficient difficulty to receive an ASD diagnosis. Therefore, males and females with lower cognitive ability may present with more similar ASD characteristics. Although the CARS2-ST form is designed for children of lower intellectual ability, in the current sample, the mean IQ of children scored on this form was within one standard deviation of the CARS2-HF and still within the *Average* range (Wechsler, 2014). Therefore, cognitive ability was also not thought to account for the lack of meaningful sex/gender differences on the CARS2-HF form.

Bitsika et al. (2019) recently demonstrated that overall ASD severity may also impact upon sex/gender differences. Given that the CARS2 forms were not designed to be compared

and have different ASD thresholds, it is not possible to comment definitively on the relative ASD severity of children scored across these forms. However, as greater ASD severity is associated with earlier diagnosis (Daniels & Mandell, 2014) and behavioural profiles analysed here were constructed at the time of diagnosis, it is possible that more severe presentations were profiled on the CARS2-ST form (designed largely for younger children). It may be, therefore, that children with more subtle ASD (scored on the CARS2-HF) present with fewer sex/gender differences. Moreover, diagnosticians may have selected the CARS2-HF form over the CARS2-ST to supplement their assessment for children with a more subtle or *female* presentation, regardless of their true sex/gender, thus homogenising this group. While this may be considered a limitation of the present study, it highlights additional challenges in understanding sex/gender differences as the selection of instruments may be a confounding variable. Having said this, the general consistency in observed sex/gender differences between the CARS2-ST and GARS-3 results suggest the sexes/genders differences seen here were real and meaningful.

Diagnostician selection of the CARS2-ST form for certain presentations may also explain why there was no meaningful difference in overall ASD severity level for children scored on the CARS2-HF and GARS-3, despite strong evidence of greater atypicality amongst males on the CARS2-ST ($d = 0.79$) and male over-representation in the *Severe* ASD category. It is also possible that autistic males of this age genuinely present with more severe features than females. Consistent with this conclusion, higher atypicality was found among males in CARS2-ST Item 15: *General impressions*, indicating that diagnosticians' subjective judgements of males were of more severe autism than females. This is consistent with other studies which have shown that present with less overt autistic behaviour during assessment (e.g., measured on the ADOS-2; Rynkiewicz & Łucka, 2015).

Reasons for Sex/gender Differences

There are a number of possible explanations for the pattern of sex/gender differences in severity and presence of atypicality across the GARS-3 and CARS2-ST. Firstly, some evidence suggests that females require greater genetic liability in order to express ASD-related difficulties compared to males and that this may be particularly the case for RRBIs (Robinson et al., 2016; Szatmari et al., 2012). If borne out, this possibility may contribute to lower severity levels amongst females in the RRBI domain (Tillmann et al., 2018) and, as this study suggests, in subdomains (e.g., use of objects) and specific behaviours which comprise the RRBI domains. Secondly, sex/gender differences in ASD presentations may be influenced by the unequal distribution of ASD traits in the typically developing population (Constantino & Todd, 2003). It remains unclear at what developmental stages and in what specific areas typically developing males demonstrate greater ASD traits, and why this may be the case (Øien et al., 2017). Thirdly, camouflaging of ASD traits may contribute to the pattern of results found here. If, as some previous findings have shown, females show greater ability to engage in “neurotypical passing” (i.e., 'passing' as neurotypical; Cage & Troxell-Witman, 2019, p. 3), differences may emerge in features that are most amenable to camouflage or those that an affected individual most actively tries to conceal. This may result in less observed difficulty in these areas among camouflaging females compared to less camouflaged males. Females’ relative skill in compensatory behaviours such as imitation may also facilitate more successful camouflaging.

Fourthly, parents, and indeed clinicians, may differ in their understanding and expectations of typical behaviour for boys and girls, and therefore how they interpret and report ASD symptoms. Bias in parent reporting and diagnostician decision-making remains relatively unexplored. However, there is some evidence that a child’s sex/gender may affect parents’ impression of aspects of an autistic child’s long term outcomes (Geelhand et al.,

2019) and that, despite ASD severity consistent with males, clinicians may be less likely to consider criteria met for females (Wilson et al., 2016). In this study, there was general agreement between the results of the CARS2-ST (completed by the diagnostician) and GARS-3 (completed by the parent or caregiver), due perhaps to parent report significantly contributing towards CARS2 profiling. However, differences between these sets of results, while likely driven in part by the greater detail of behaviours on the GARS-3, may also have been a function of the inclusion of clinician observation, teacher reported information and interpretation of parent report for the CARS2.

A final explanation for females' lower scores in several specific behaviours is that ASD symptoms may be expressed differently among females compared to males and in ways that are not adequately captured by some instruments, including those used in the current study. For example, other authors have suggested that rather than being less impaired in the area of restricted interests, females' interests may be oriented differently, such that they are less *overtly* atypical (Attwood et al., 2006). Reporters may be primed to consider typical autistic obsessions (e.g., trains and timetables) when responding to questions such as GARS-3 Item 50: *Shows an intense, obsessive interest in specific intellectual subjects*, rather than interests that may be more typical of the female presentation (e.g., popstars). In addition, ASD features more characteristic of, and unique to females (e.g., compulsive or self-injurious behaviour; Antezana et al., 2018) may be absent from instruments such as the GARS-3. As such, qualitative differences in the expression of ASD characteristics, even at the level of specific behaviours, may explain lower scores amongst females. These sex/gender differences may be best studied using a bottom-up approach rather than relying on the structure and definitions imposed by existing instruments. Based on the results of this study, it is difficult to conclude whether sex/gender bias exists in these instruments. This is primarily due to the role of the clinician in the selection of instruments and for the GARS-3,

similar overall scores between males and females despite a large number of items where greater atypicality was found among males. However, given that a large number of individual items were skewed towards showing greater difficulty among males and only one showed this among females, it may be deduced that these instruments may be more sensitive to the ASD difficulties of males than of females.

Strengths and Limitations

As the data analysed in the current study was extracted from pre-existing assessment reports produced by diagnosticians at a single private clinic in Adelaide, the generalisability of these findings and the effects of any selection bias (resulting from cultural or socioeconomic circumstances of clients) is unknown. Comparing these findings with the wider literature is complicated by the structure of the CARS2 items, as some items incorporate features from across different DSM-5 criteria. However, the results of this study provide insight into which clusters of ASD characteristics may best differentiate male and female ASD presentations. In addition, if as this and other studies have suggested, there are presentations of ASD that fit imperfectly with the current ASD conceptualisation and assessment instruments, those who most embody these presentations may remain undiagnosed and have largely been excluded from research. In this study, I was unable to include children assessed but not diagnosed with ASD due to limitations in data availability. As a result, no comment can be made about any association between symptom profiles and the probability of ASD diagnosis. Studying presentations of individuals with subclinical ASD, or with many ASD features who remain undiagnosed, may prove fruitful in understanding whether the ASD conceptualisation and assessment tools should be modified and whether sex/gender differences generalise to below the diagnostic threshold. This approach is taken in Study 2, Chapter 4.

Strengths of the present study include the sophisticated statistical methods employed, large sample size, and the use of empirically validated ASD instruments. On the other hand, the use of existing and preselected assessment instruments restricts this and other research to the existing, possibly androcentric conceptualisation of ASD. Furthermore, the use of data from a single clinic and lack of information about co-occurring psychiatric conditions restrict the generalisability of the current findings.

Conclusions and Future Directions

Despite these limitations, this study provides insight into how ASD presentations may differ between diagnosed male and female children. Males demonstrated greater atypicality in a number of specific ASD behaviours and subdomains (i.e., groups of behaviours) according to the GARS-3 and CARS2-ST. No meaningful differences were found on the CARS2-HF form, suggesting that age and the instrument selected by diagnosticians may affect resultant sex/gender differences.

Sex/gender differences identified in this study, both in symptom type and severity, may render the female presentation of ASD less recognisable to referrers, such as parents and teachers, and clinicians tasked with assessment. If the female presentation is less recognisable, ASD diagnosis may be delayed or overlooked entirely (Mandy et al., 2011). In this study, I identified a number of areas in which females may be more likely to present as typically developing, which may further compound under-detection and mean that the broader constellation of ASD difficulties are overlooked. Given the above, it is critical that diagnosticians, referring clinicians, and teachers are educated in these phenotypic differences so that females' ASD may be detected in a timely manner.

The analysis of sex/gender differences in specific behaviours defined according to two psychometric instruments was useful in illustrating where the ASD-related difficulties of females may differ from those of males. However, this study showed that such instruments

may lack sensitivity to the difficulties of females, perhaps as a result of being constructed from androcentric literature. This insensitivity is an issue for two reasons. First, if the operationalisation of difficulties is not consistent with the presentation of females, it will be difficult to establish the precise nature of females' difficulties by analysing scores on these instruments. Second, females whose ASD presentation deviates from the classic male presentation may not receive ASD diagnoses, resulting in these individuals being excluded from much of the research to date. For these reasons, the presentations of females narrowly below the ASD diagnostic radar are considered in Chapter 4. The relative association of specific difficulties with diagnostic outcome and sex/gender will also be explored.

Chapter 4: Studies 2a and 2b

Under the Diagnostic Threshold: Why Do Some Females Fail to Meet ASD Criteria Despite Being Suspected of Having ASD?

Overview

Obtaining an ASD diagnosis is often a more difficult process for females than males, meaning that many autistic females may "fly under the diagnostic radar" (Lai et al., 2011, p. 6). Several hypotheses have been put forward to explain why this may occur and have garnered considerable support. Specifically, females may be less likely to be referred for specialist assessment (Øien et al., 2018) or to be considered as having ASD even if the severity of their symptoms is the same as that of males (Giarelli et al., 2010; Lai, Lombardo, et al., 2015; Russell et al., 2011). This may be due to clinicians overlooking or normalising symptoms, or a tendency to consider other diagnoses first and otherwise explain the presenting difficulties. Compounding this, females may present with ASD difficulties that differ quantitatively and qualitatively from those of males, rendering the androcentric conceptualisation of ASD an imperfect fit with their presenting concerns (Kreiser & White, 2014). While it is informative to consider sex/gender differences in how ASD presents amongst diagnosed individuals, their difficulties must fit neatly within the disorder conceptualisation in order for them to have received diagnoses. Therefore, it is likely that among those diagnosed with ASD, phenotypic sex/gender differences are minimised, and that females whose presentation deviates more, remain undiagnosed. By virtue of this, these females remain largely excluded from current research and their difficulties remain poorly understood. The purpose of this study is therefore to examine why females with many autistic traits may fail to meet ASD diagnostic criteria. It also aims to explore their difficulties, how these compare to those who receive ASD diagnoses, and the associations between particular ASD behaviours and diagnostic outcome. In doing so, the presentations of children who

receive ASD diagnoses at a second formal assessment after an initial negative result, and those of children who do and do not meet ASD criteria at a single assessment, will be examined.

To date, the vast majority of literature investigating the possibility of a distinctive female ASD presentation has followed the convenient path of recruiting individuals with ASD diagnoses as participants. While this may be considered advantageous in other research, it presents a problem of circularity in this line of enquiry. Specifically, if indeed the ASD assessment instruments and conceptualisation are biased toward the typical male presentation (Kreiser & White, 2014; Rutter et al., 2003), only females whose presentation is sufficiently congruent with the male conceptualisation will receive ASD diagnoses. In other words, females with marked difficulties which deviate from males' difficulties, either in severity or type, may fail to meet diagnostic criteria (Bargiela et al., 2016). It remains unknown whether they fail to meet criteria because they are indeed not autistic (and thus should not be included in studies of the female ASD phenotype), or whether the criteria and diagnostic instruments are insensitive to the way their ASD is expressed (Lai, Lombardo, et al., 2015). That is, as a result of being derived from androcentric literature, the current notions of, and instruments to assess for ASD may be too narrow to capture the genuinely autistic difficulties of these females. Of particular interest to this study are the reasons for which females with autistic traits and who are suspected of having ASD are referred for assessment and why they may or may not be deemed to meet criteria.

What is Known About Females Under the Diagnostic Threshold?

A small number of studies has considered the presentations of individuals with sub-clinical ASD traits (i.e., individuals referred for ASD assessment and/or who achieve high scores on screening tools but are deemed ineligible for ASD diagnoses). These studies have often examined the role of factors such as intellectual ability and symptom severity upon the

probability that a male or female will receive an ASD diagnosis. Some have found evidence of interactions, such that the association between certain factors and ASD diagnosis is stronger for one sex/gender. For example, greater behavioural and emotional difficulties have been found to increase the likelihood of ASD diagnosis more for girls than for boys, suggesting that girls require more atypicality in these areas to be considered autistic (Duvekot et al., 2016; Dworzynski et al., 2012). Parent-reported repetitive and restricted behaviours and interests (RRBIs) have also been found to more strongly predict ASD diagnosis for boys than girls (Duvekot et al., 2016). Surprisingly, Duvekot et al. found no significant sex/gender difference in how strongly overall ASD severity predicted ASD diagnosis. Nonetheless, these studies suggest that there may be asymmetry in how symptom severity and other factors influence the likelihood of ASD diagnosis for males and females.

Wilson et al. (2016) examined the presentations of a large number of adults who were referred for ASD assessment and subsequently categorised as having ASD, partial ASD (i.e., atypical autism or unspecified pervasive developmental disorder) or no ASD according to the International Classification of Diseases-10th Revision (ICD-10R) criteria (World Health Organisation, 1993). Belonging to a particular subgroup was associated with the severity of social and repetitive behaviours and, consistent with other findings, males were slightly more likely than females to receive an ASD diagnosis. However, the patients' sex/gender and the source of information (i.e., parent report or observation) was also found to affect diagnostic evaluation, with the size of this effect differing by subgroup. Females with partial ASD had significantly more severe social difficulties than their male counterparts according to parent report collected via the Autism Diagnostic Interview- Revised (ADI-R). On the other hand, repetitive behaviour was more severe for males than females in each subgroup. Finally, males with ASD only showed significantly greater abnormality than autistic females in

Communication according to the Autism Diagnostic Observation Schedule- Generic (ADOS-G).

It remains unknown as to whether the difficulties of females with many ASD traits who present for assessment do not receive ASD diagnoses because (a) their difficulties do not reach the clinical significance levels associated with ASD, (b) they do not demonstrate a broad enough variety of ASD behaviours, and/or (c) they do not present with difficulties that fit within the ASD diagnostic domains. Further, although females are diagnosed later on average (Begeer et al., 2013), very little is known about how females' presentation may change over time, particularly for those who subsequently meet diagnostic criteria for ASD. Thus, it is important to examine behaviours at a fine-grained level where sex/gender differences are more likely to exist in order to fully appreciate the emergence of ASD difficulties, rather than considering characteristics at the level of broad domains (i.e., social communication and RRBIs) as has typically been done in the past (Hiller et al., 2014; Lai, Lombardo, et al., 2015).

Individual Differences

In addition to examining fine-grained behavioural manifestations of ASD, it is also important to consider other individual differences (e.g., intellectual ability and age) and the source of diagnostic information (i.e., from parent or teacher report, or diagnostic observations). Each of these factors has been found to affect the relationship of sex/gender on the severity and manifestation of ASD behaviours/characteristics.

Intellectual ability may significantly influence the likelihood that a child will receive an ASD diagnosis, with those with lower intellectual ability more likely to be diagnosed (Rivet & Matson, 2011a; van Wijngaarden-Cremers et al., 2014). This is reflected in the asymmetrical female/male prevalence ratios for people with and without intellectual disability (1:2 and 9-10:1 respectively; Banach et al., 2009; Loomes et al., 2017). Given the

confounding effect of low cognitive ability on ASD diagnosis, children with intellectual disability were excluded from the current study.

Sex/gender differences in certain ASD symptom domains have been found to vary according to the age of those recruited. In particular, sex/gender differences in RRBI severity may be most evident during adolescence for children diagnosed with ASD (van Wijngaarden-Cremers et al., 2014). It has also been suggested that females' social difficulties may be more pronounced later in childhood, when social demands outweigh social skills (Hsiao et al., 2013; Mandy et al., 2018). However, it remains unknown as to whether particular ASD behaviours change in severity or number over time such that a diagnosis can ultimately be made and what specific concerns prompt parents to arrange reassessment.

Source of Diagnostic Information

As noted in Study 3 (Chapter 5) and the results of previous investigations (e.g., Hiller et al., 2014; Mandy et al., 2011; Mayes & Lockridge, 2018), there may be significant discrepancies in the diagnostic information provided from different sources. Specifically, parents may raise significant concern that is not shared by teachers (or vice versa). It is now understood that ASD may present inconsistently across different environments and, therefore, discrepancies in what is reported may be expected. These discrepancies may be larger for females because of a proclivity for camouflaging or masking ASD-related difficulties in social environments (Hull et al., 2020). As a result, teachers' concerns may be less severe for autistic girls than boys (Hiller et al., 2014). Social camouflaging may also influence the ASD behaviours observed by a diagnostician at the time of assessment and result in discrepancy between diagnostic observations and parent report (Mayes & Lockridge, 2018). Although diagnosticians surveyed in Study 3 report that it is challenging to reconcile contrasting information from different sources, the relative influence of each report on diagnostic outcomes for males and females remains unexplored.

In light of the above gaps in our knowledge, the present study will examine the ASD presentations of children narrowly below the ASD diagnostic threshold; specifically, those who were suspected of having ASD and were found to have many ASD traits but who did not fully meet criteria at assessment. A sample of females (n = 12) who had initially received a negative diagnosis of ASD and returned for reassessment, ultimately meeting ASD criteria, will be examined in Part A of this study, with a larger (separate) sample of children assessed and diagnosed/not diagnosed at a single assessment considered in Part B (total N = 222 children; n = 98 females). Fine-grained data regarding particular groups of behaviours were extracted from the parent-report and teacher-report data, and diagnostic observations documented in the children's ASD assessment reports. The following research questions were posed:

Part A

1. Why were ASD diagnoses not made at Assessment 1? Was this the same for males and females?
2. Why did the children present for reassessment? Was this the same for both males and females?

3a. How do the criteria met by males and females at Assessment 1 compare with Assessment 2, at which ASD diagnosis was made?

3b. How did the specific behavioural presentations of female clients change between Assessments 1 and 2, such that an ASD diagnosis could ultimately be made?

Part B

1. Which criteria did females commonly fail to meet, and what was the nature of their difficulties in these areas (if any)?
2. What were the differences in the ASD-related features of females with ASD and females under the threshold (non-ASD)? Were these the same for males?

3. Which behaviours (reported from which sources) predicted ASD diagnosis, and was this the same for males and females?

Study 2a

Method

Participants

Data were extracted from the ASD diagnostic assessment reports of 37 children (females: $n = 12$; males: $n = 25$) who had undertaken two ASD assessments and received a diagnosis of ASD only at the second assessment. Data from the initial assessment report (negative ASD result) and second assessment report (positive ASD result) were extracted.

Participants received their ASD diagnosis at a large private clinic specialising in developmental disorders and were then referred and registered with the local autism organisation who verified the diagnosis from the reports provided. Many of these children (75% of females, $n = 9$; and 64% of males, $n = 16$) undertook their initial assessment through other service providers, and the remainder presented to the same clinic for reassessment. All diagnosticians (psychologists, speech pathologists, psychiatrists, and paediatricians) had specialist training in assessment for ASD and were certified as diagnosticians by the local autism association. Assessments were conducted by one, two, or multiple diagnosticians, depending on the clients' needs and the protocols of the organisation. As part of this clinic's standard protocol, informed consent was provided by all parents (and children if appropriate) for data contained in their diagnostic assessments to be used for research purposes. Ethical approval was granted by the author's tertiary institution.

Assessment Processes

For all children whose diagnostic data were included in the analysis, diagnosticians referred to the DSM-5 ASD criteria (American Psychiatric Association, 2013) to inform their

assessment decision-making (i.e., whether ASD criteria were met). In the assessment reports, diagnosticians rated the child's presentation as either fully, partially, or not meeting each criterion. Partially meeting a criterion indicated that, although the child presented with some ASD features in this domain, these features were considered to be insufficient in severity or number for the criterion to be met, and hence an ASD diagnosis was not given.

For all assessments, the information from a detailed interview with a parent/caregiver and diagnostic observations formed a large component of the diagnostic report. However, as the presence of autistic difficulties across different settings is necessary for ASD diagnosis, observational reports from teachers and other professionals were often sourced and included as evidence towards the diagnostic outcome. As reports were issued by different diagnosticians and organisations, a range of standardised assessment tools was used to supplement assessments. The selection of these tools was based on their availability, clinician preference, and the age and cognitive ability of the child. They included the Autism Diagnostic Interview- Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994), Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2; Lord et al., 2012), Autism Detection in Early Childhood (ADEC; Young, 2007), Childhood Autism Rating Scale, 2nd Edition (CARS2; Scholper et al., 2010), Social Responsiveness Scale, 2nd Edition (SRS-2; Constantino, 2011), and Gilliam Autism Rating Scale, 3rd Edition (GARS-3; Gilliam, 2014).[28]

[28] Children's total scores on these tools were collected and analysed, but given the variety of assessment tools used, there were insufficient observations to make meaningful conclusions for any of the measures. Therefore, these results are not presented.

Procedure

Eligible children were identified after examining diagnostic assessment reports of clients assessed for ASD between September 2013 and March 2019.[29] Sixty-six children were identified as having undertaken two ASD assessments and only receiving an ASD diagnosis on the second occasion. Due to the unavailability of their initial assessment report, 29 children (males: $n = 22$, females: $n = 7$) were excluded, leaving a total of 37.

The extent to which a child met each ASD criterion at the initial and second assessment was documented as (0) *not met,* (1) *partially met,* or (2) *met*. Qualitative information from the assessing diagnostician(s) highlighting the reasons for which an ASD diagnosis was not made at the initial assessment and parent-reported reasons for reassessment were also extracted verbatim. As the present study was primarily concerned with the reasons that females subsequently diagnosed with ASD may fail to meet diagnostic criteria, fine-grained information about the specific behaviours of concern at Assessments 1 and 2 was collected for female children. This included information provided by parents/guardians in a diagnostic interview, diagnosticians' observations, and teachers' report regarding the child's behaviour in the school environment (available for 66.7% of females, $n = 8$, at each assessment). Diagnostic observations were reported in an unstructured manner within each criterion or as part of the documentation of findings using a structured assessment instrument (e.g., Autism Diagnostic Observation Schedule; Lord et al., 2012; or Autism Detection in Early Childhood; Young, 2007). The child's current teacher's concerns were extracted from their responses on a structured questionnaire provided to teachers as part of the clinic's standard assessment procedure. Due to inconsistencies in the thoroughness of reporting across each source, the ASD behaviours (variables) examined differed slightly across parent,

[29] As noted in Study 1 (Chapter 3), data latency was not anticipated to meaningfully affect the quality of these data.

observation, and teacher accounts. However, all were driven by the DSM-5 criteria for ASD diagnosis, and coded as (0) *no concern,* (1) *some/partial concern,* (2) *concern.* Given findings that the specific stereotypical behaviours and restricted interests of autistic females may differ from those of males (Antezana et al., 2018; McFayden et al., 2018), the presence of particular mannerisms and interests was documented as reported by their parents and coded as (0) *absent* or (1) *present*.

The variables analysed are outlined in the results section.[30] To test inter-rater reliability, a research assistant re-coded 10% of the data after the initial data extraction and coding. Adequate inter-rater reliability was established (Cohen's kappa scores between 0.75 and 1.00 across variables). I re-coded 20% of the data and Cohen's kappa scores were between 0.80 and 1.00, indicating adequate intra-rater reliability.

Data Analysis

Bayesian logistic regressions were used for each of the major analyses.[31] To investigate how females' presentations differed between assessments such that an ASD diagnosis could ultimately be made and to establish how strongly assessment number predicted the presence of a particular behaviour, the behaviours were set as the outcome variables in each analysis. These were coded either in binary, (0) *absent* or (1) *present*, or tertiary form, (0) *no concern,* (1) *partial/some concern*, or (2) *concern.* Standard logistic regression was used for the binary outcome variables and conditional logistic regression for the tertiary outcome variables. Assessment number was entered as a repeated predictor variable. An identical approach was used to test the role of sex/gender (between groups) and

[30] Refer to Appendix E for a description of how each variable was operationalised and coded.

[31] Mathematical models for analyses in Study 2 (Parts A and B) are presented in Appendix C.

assessment number (repeated measure) on the extent to which ASD criteria were met (tertiary outcome variable).[32]

Log odds ratios (LORs), the preferred effect size measure for categorical variables, are reported for each of the major analyses and they provide the basis for all inferential decisions. LORs are symmetrical around zero, with more extreme deviations from zero indicating a stronger association between predictor and outcome variables in a specified direction. Here, where sex/gender is defined as a predictor variable, male sex/gender is assigned the positive direction and female sex/gender is assigned the negative direction. Thus, positive log odds ratios indicate that males had a higher probability of presenting with a particular behaviour. Where assessment number is the predictor variable of interest, Assessment 2 (at which ASD was diagnosed) is assigned the positive direction. As LORs are not always intuitive to interpret, relative risks were calculated from the model-estimated proportions and reported. Relative risk reflects how many times more probable the criterion outcome is in one condition (versus another). Thus, a relative risk of 2 for males (versus females) reflects the outcome being twice as likely for males than females.

The individual effects of each predictor variable on the presence of the ASD behaviour were analysed first. For analyses in which two predictor variables were included (i.e., both assessment number and sex/gender), interaction terms are also included. A meaningful interaction suggests that the effect of assessment number on the presence of the behaviour was not consistent between males and females.

HDIs $_{(80\%)}$ are reported for each LOR and reflect the upper and lower bounds of the most credible values (with 80% confidence). ROPEs of ± 0.1 were applied in the

[32] Given the small sample size and the repeated nature of these data, statistically controlling for cognitive ability (theoretically, a stable construct) or age (as change in presentation with age was of interest in this study) was deemed inappropriate.

interpretation of the $HDI_{80\%}$ and to quantify the probability of a meaningful effect in a given direction.

Participant Demographic Information

Age at Assessments

A Bayesian *t*-test revealed moderate evidence of a meaningful difference in age between males (M = 6.94 years, $HDI_{80\%}$ = [5.97, 7.83], SD = 3.43) and females (M = 5.62 years, $HDI_{80\%}$ = [4.68, 6.64], SD = 2.58) at Assessment 1 (d = 0.43, $HDI_{80\%}$ = [-0.01, 0.89]). There was an 82.6% probability that males were meaningfully older ($P_{(meaningful)}$ = 82.6%). A similar pattern emerged between males and females at Assessment 2 (males: M = 8.98 years, $HDI_{80\%}$ = [8.06, 9.96], SD = 3.52), females: M = 7.54 years, $HDI_{80\%}$ = [6.40, 8.70], SD = 3.00). Although we cannot exclude zero or a negligible difference with 80% confidence ($P_{(within\ ROPE)}$ = 10.1%), there was moderate evidence that males were meaningfully older ($P_{(meaningful)}$ = 83.1%; d = 0.45, $HDI_{80\%}$ = [-0.00, 0.94]).[33]

Cognitive Ability

Standardised information regarding cognitive ability was available for 75.0% of males and 83.3% of females. This information had been collected at the time of either assessment and was collected using either the Wechsler Intelligence Scale for Children (Wechsler, 2003; 2014) or Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 2002; 2012). In the event that only certain subtest or index scores were available, these were prorated to establish proxy Full Scale Intelligence Quotients (FSIQs; 10.8% of cases). Results of a Bayesian *t*-test, conducted on the subsample for whom data were available, revealed equivocal evidence regarding a meaningful difference in FSIQ between males (M = 98.09,

[33] The potential impact of these age differences is examined in the Discussion.

$HDI_{80\%}$ = [91.83, 104.34], *SD* = 12.37) and females (*M* = 93.60, $HDI_{80\%}$ = [80.13, 107.16], *SD* = 19.70; *d* = 0.29, $HDI_{80\%}$ = [-0.56, 1.20], $P_{(meaningful)}$ = 66.3%).

Family History

Information regarding any family history of ASD was available for 81.1% of children (males: 76.0%, *n* = 19; females: 91.7%, *n* = 11). Thirty-six percent of females (*n* = 4) and 26.3% of males (*n* = 5) had no known family history of ASD. Logistic regression revealed no evidence of a meaningful difference in the probability of having a family history of ASD (LOR = 0.47, $HDI_{80\%}$ = [-0.58, 1.55], ($P_{(meaningfully\ higher\ in\ males)}$ = 67.2%). Over half of the females (54.5%, *n* = 6) had an immediate family member with a diagnosis of ASD (i.e., a sibling or parent), compared to 31.6% of males (*n* = 6). The remaining 42.1% of males (*n* = 8) and 9.1% of females (*n* = 1) had a non-immediate family member diagnosed with ASD (e.g., cousin).

Referral Pathway

The relationship of the referrer to the child was categorised as being the child's family, a health professional (e.g., medical doctor, psychologist, speech pathologist) or the child's school. At Assessment 1, 72.7% of females were referred by health professionals, with the remainder being referred by family (27.3%). Similarly, 84.0% of males were referred by health professionals, 12.0% by family, and 4.0% by their school for an initial assessment. The majority of females were referred for reassessment by family (63.6%), with a minority being referred by health professionals (36.4%) or school (9.1%). Approximately half of males (48.0%) were referred for reassessment by their family, 40.0% by health professionals, and 12.0% by the school.

Broadly, there was a higher probability that children would be referred by a professional or the school for Assessment 1 compared to Assessment 2 (LOR = -1.96, $HDI_{80\%}$ = [-2.87, -0.99], $P_{(meaningful)}$ = 99.8%), where more were referred by the child's family.

Male children had a higher probability of being referred by the school or a health professional compared to females (LOR = 0.75, $HDI_{80\%}$ = -0.30, 1.77], $P_{(meaningful)}$ = 79.9%). Here, a negligible difference could not be excluded with 80% confidence.

Diagnosis Prior to ASD Assessment 1

Information about prior diagnoses was available for 25 children (18 males and 7 females). For two females (28.6%) and five males (27.8%), previous psychiatric diagnoses were documented. There was no evidence of a difference in the probability that a male or female had a previous diagnosis (LOR = 0.02, $HDI_{80\%}$ = [-1.34, 1.32]). Similarly, there was no evidence of a difference in the likelihood that any single diagnosis was given.

Prior Allied Health Support

Information about allied health support accessed prior to Assessment 1 was available for 21 males (84.0%) and nine females (75.0%). There was moderate evidence that males had a higher probability of having accessed psychological support than females (males = 47.6%, females = 22.2%; LOR = 1.16, $HDI_{80\%}$ = [-0.06, 2.36], $P_{(meaningful)}$ = 88.3%). However, there was no meaningful sex/gender difference in prior access to speech pathology (males = 42.9%, females = 44.4%; LOR = -0.14, $HDI_{80\%}$ = [-1.20, 0.90]) or occupational therapy (males = 40.7%, females = 44.0%; LOR = -0.14, $HDI_{80\%}$ = [-1.17, 0.93]).

Information about allied health support accessed after Assessment 1, but prior to Assessment 2, was available for 35 children: 24 males (100%) and 11 females (91.7%). No meaningful sex/gender difference was found in the probability of having accessed psychological support (males = 62.8%, females = 64.4%; LOR = -0.07, $HDI_{80\%}$ = [-1.06, 0.93]) or speech pathology (males = 45.8%, females = 54.7%; LOR = -0.36, $HDI_{80\%}$ = [-1.30, 0.60). However, females had a meaningfully higher probability of having accessed occupational therapy prior to Assessment 2 (males = 45.9%, females = 73.7%; LOR = -1.20, $HDI_{80\%}$ = [-2.24, -0.17], $P_{(meaningful)}$ = 92.4%).

Results

Why ASD Diagnoses Were Not Provided at Assessment 1

Justification Given by Diagnosticians

Five broad reasons were offered by diagnosticians in their assessment reports as to why an ASD diagnosis was not provided at Assessment 1.[34] For males and females, a lack of sufficient ASD-related atypicality was the most common. However, while males in this sample were almost as likely to have inadequate atypicality in the social communication domain (n = 18, 75.0%) as in the RRBI domain (n = 19, 79.2%), among females, the most common reason was a lack of sufficient atypicality in social communication (n = 10, 83.3%). Bayesian logistic regression analysis revealed no evidence of meaningful differences in the probability that any one reason was given for males more often than females (i.e., a negligibly sized difference in probability could not be ruled out with 80% confidence in each case). However, there was moderate evidence to suggest that males had a higher probability of not receiving an ASD diagnosis because their presentation was better explained by an alternate diagnosis, such as post-traumatic stress disorder or attention-deficit/hyperactivity disorder (1.6 times higher probability for males; LOR = 0.81, $HDI_{80\%}$ = [0.05, 1.76], $P_{(meaningful)}$ = 83.4%) or due to having inadequate atypicality in the RRBI domain (1.4 times higher probability; LOR = 0.77, $HDI_{80\%}$ = [-0.28, 1.83], $P_{(meaningful)}$ = 79.2%). Inadequate atypicality in the social communication domain was the only reason for which there was evidence (albeit weak) of a higher probability of being given for females (males = 75.0%, females = 83.3%; LOR = -0.60, $HDI_{80\%}$ = [-1.75, 0.66], $P_{(meaningful)}$ = 70.6%). There was equivocal evidence of a difference in the probability that a male or female would not meet

[34] For one of these reasons, *Insufficient information available,* the $HDI_{80\%}$ was extremely wide due to the low frequency with which this reason was cited (n = 1). As no conclusions could be drawn about this variable, it was excluded from the results.

criteria for ASD diagnosis because of inconsistency in their presentation across settings (males = 4.2%, females = 8.3%; LOR = -0.71, $HDI_{80\%}$ = [-2.80, 1.33], $P_{(meaningful)}$ = 65.1%).

Despite not providing complete ASD diagnoses, explicit acknowledgement and recognition of autistic traits was 1.4 times more likely to be documented by diagnosticians for males (83.3%), compared to females (58.3%; LOR = 1.30, $HDI_{80\%}$ = [0.24, 2.35], $P_{(meaningful)}$ = 93.2%). Despite this acknowledgement, a review assessment for ASD was only explicitly recommended in the reports of nine males (36.0%) and three females (25.0%).

Criteria Met at Assessment 1

In order to qualify for a diagnosis of ASD under the DSM-5 criteria (American Psychiatric Association, 2013), an individual must meet all three criteria in Domain A (social communication) and at least two of the four criteria in Domain B (repetitive behaviour). Table 4.1 indicates the relative probabilities of meeting each criterion for males and females at their initial assessment. The table presents log odds ratios for the probability that a male or female would meet a given criterion, with probabilities that these differences were meaningful and the percentage of children of each sex/gender who met each criterion. There was moderate evidence that females in this sample had a higher probability of meeting Criterion B1 (stereotyped/repetitive movement, speech, or object use) compared to males (2.2 times greater probability for females). However, an effect in the opposite direction (higher probability among males) cannot be excluded with 80% confidence. On the other hand, there was strong evidence that males were more likely than females to meet Criterion B2 (insistence on sameness, routines, and rituals; 3.8 times greater probability for males). Females were 1.7 times more likely to meet criteria in the RRBI domain (i.e., two or more of the four Domain B criteria), despite not displaying enough social difficulty to meet all ASD criteria. However, the evidence of a meaningful effect in this direction was weak. Criteria A3 (difficulties with relationships) and B4 (hyper-/hypo-reactivity to sensory input) were the

most commonly met criteria for males and females at Assessment 1, suggesting concern here may be important in why a referral was made.

Table 4.1

Logistic Regression Effect of Sex/Gender on Criteria Met at Assessment 1

Criterion	Log Odds Ratio [$HDI_{80\%}$]	$P_{(meaningful)}$	Proportion Met (%)	
			Male	Female
Domain A	0.01 [-1.24, 1.29]	.46	15.1	14.9
A1	0.42 [-0.55, 1.41]	.66	45.8	35.7
A2	-0.33 [-1.47, 0.75]	-.61	20.1	25.9
A3	0.36 [-0.60, 1.30,	.64	54.3	45.3
Domain B	-0.69 [-1.77, 0.32,	-.77	19.3	32.2
B1	-1.06 [-2.18, 0.02]	-.87	16.0	35.0
B2	**1.60 [0.10, 3.16]**	**.92**	28.3	7.5
B3	0.23 [-1.02, 1.46]	.56	20.0	16.6
B4	-0.19 [-1.15, 0.75]	-.55	50.1	54.8

Note. Criterion A1: deficits[35] in social-emotional reciprocity; A2: deficits in nonverbal communication behaviours; A3: deficits in developing, maintaining, and understanding relationships; B1: stereotyped/ repetitive motor movements, use of objects, or speech; B2: insistence on sameness, routines, or ritualised behaviour; B3: restricted and fixated interests; B4: hyper-/hypo-reactivity to sensory input.

Positive log odds ratio = greater probability of criterion being met for males. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction (positive probabilities indicate a higher probability of meeting criterion for females; negative probabilities indicate a higher probability for males).

[35] The word 'deficit' is used here in accordance with the language used in the DSM-5 (American Psychiatric Association, 2013). Where possible, I have used less pathologising language to refer to autistic behaviours or difficulties.

Other Diagnoses Provided

More than half of females were not provided with any diagnosis at their first assessment (58.3%, $n = 7$) compared to 52.0% of males ($n = 13$). One male and one female was diagnosed with a specific learning disorder (4.0% and 8.3%, respectively). The ratio of specific learning disorder diagnoses between males and females was identical for diagnoses of social (pragmatic) communication disorders, which are often considered *partial* ASD diagnoses, with only social communication criteria being met. Two females (16.7%) and three males (12.0%) were diagnosed with language delay/disorders, and developmental delay was diagnosed for 12.0% of males and 8.3% of females. Finally, 4.0% of males ($n = 1$) were each diagnosed with attention-deficit/hyperactivity disorder, speech delay, oppositional defiant disorder, or reactive attachment disorder. None of these diagnoses were provided to any female in this sample at her first assessment. Given the small numbers of children meeting criteria for each of these diagnoses at their first assessment, logistic regression analyses revealed a large degree of uncertainty as to whether they were more probable among males or females and therefore no conclusions could be drawn.

Reasons for Reassessment

Parents' concerns, offered as reasons for pursuing reassessment for their children, were extracted verbatim and categorised. These data were available for 12 females (100%) and 22 males (88%; total $n = 34$). Bayesian logistic regression analysis was used to assess sex/gender differences in the probability that the child's parent would cite each of eight particular concerns (listed below) as a reason for reassessment. Analysis revealed strong evidence of greater probability that four reasons would be given by parents of males than females. In particular, difficulty *Managing change* (reported for 27.3% of males, 0.0% of females; LOR = 4.53, $HDI_{80\%} = [1.13, 7.76]$, $P_{(meaningful)} = 99.4\%$) and *Difficulties with concentration/academic achievement* (males = 22.7%, females = 0.0%; LOR = 4.20, $HDI_{80\%}$

= [0.88, 7.46], $P_{(meaningful)}$ = 98.8%) were cited more frequently for males. Parents of males were 4.9 times more likely to report *Anxiety* as a primary reason for reassessment (males = 40.9%, females = 8.3%, LOR = 2.19; $HDI_{80\%}$ = [0.70, 3.69], $P_{(meaningful)}$ = 99.8%) and 3.0 times more likely to cite *Externalising behaviour* (males = 50.0%, females = 16.7%; LOR = 1.70, $HDI_{80\%}$ = [0.52, 2.84], $P_{(meaningful)}$ = 97.2%). Weaker evidence was found to suggest that males had a 2.3 times higher probability of being referred due to *Sensory difficulties* compared to females (males = 36.4%, females = 16.2%; LOR = 1.14, $HDI_{80\%}$ = [-0.06, 2.31], $P_{(meaningful)}$ = 88.5%). The only reason found to have a higher probability of being reported for females than males was *General concern,* or unspecified and/or widespread concern across a number of areas (2.8 times more likely for females; LOR = -1.24, $HDI_{80\%}$ = [-2.59 0.09], $P_{(meaningful)}$ = 87.0%). No meaningful sex/gender difference was found in the probability that difficulty with *Social skills* (males = 31.8%, females = 33.0%; LOR = -0.05, $HDI_{80\%}$ = [-1.05, 0.95]) or *Emotional regulation* (males = 31.8%, females = 25.0%; LOR = 0.38, $HDI_{80\%}$ = [-0.71, 1.41]) were specifically cited by parents.[36]

Changes in Presentation to Allow for ASD Diagnosis at Assessment 2

Levels of Concern Raised from Each Source Across Assessments

Bayesian hierarchical conditional logistic regressions were used to investigate the effect of assessment number (repeated predictor variable) on the probability of each level of concern (*no concern*, *partial concern,* and *concern*; categorical outcome variable). The data were analysed separately by report source (i.e., parent, diagnostician, and teacher) and ASD domain (i.e., social communication and RRBI). Table 4.2 presents the results of the

[36] Ten additional areas of concern were cited infrequently (i.e., *ns* = 1-3) and therefore $HDIs_{(80\%)}$ were too large for comparative conclusions to be drawn. These areas were concerns regarding speech/language (*n* = 3), developmental delay (*n* = 3), suicidality (*n* = 2), theory of mind (*n* = 2), restricted interests (*n* = 2), repetitive behaviour (*n* = 2), rigid thinking (*n* = 1), adaptive skills (*n* = 1), sexualised behaviour (*n* = 1) and hyperactivity (*n* = 1).

regressions with log odds ratios (LORs) as the effect size statistic, calculated based on converting the three-level proportion estimates into two binary contrasts (concern vs partial/no concern; and no concern vs concern/partial concern). The probability that the LOR HDIs $_{(80\%)}$ fell outside of the ROPE are presented in the $P_{(meaningful)}$ columns.[37]

Table 4.2

Difference in the Probability of Concern Levels Between Assessment 1 and 2

Source	Domain	Concern (vs partial or no concern)		No concern (vs concern or partial concern)	
		LOR [$HDI_{80\%}$]	$P_{(meaningful)}$	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Parent	SC	**1.21 [0.92, 1.50]**	**1.00**	**-1.45 [-1.74, -1.14]**	**-1.00**
	RRBI	**0.94 [0.62, 1.28]**	**1.00**	**-0.78 [-1.06, -0.47]**	**-1.00**
Diagnostic observations	SC	**1.88 [1.44, 2.29]**	**1.00**	**-1.41 [-1.76, -1.05]**	**-1.00**
	RRBI	**1.33 [0.63, 2.05]**	**.99**	-0.47 [-0.88, -0.06]	-.87
Teacher	SC	**1.00 [0.28, 1.67]**	**.96**	**-0.69 [-1.24, -0.16]**	**-.93**
	RRBI	0.56 [-0.13, 1.25]	.82	-0.34 [-0.83, 0.14]	-.74

Note. Positive LOR = greater frequency at Assessment 2, negative LOR = greater frequency at Assessment 1. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. SC = social communication, RRBI = repetitive and restricted behaviours and interests, LOR = log odds ratio.

There was unequivocal evidence of a higher probability that *concern* in the social communication domain was raised by parents, diagnosticians, and teachers at Assessment 2 compared to Assessment 1 and, accordingly, that the probability the *no concern* would be raised decreased meaningfully by Assessment 2. In the social domain, according to each source, the relative sizes of the LORs indicate that the changes in probability of *concern* were

[37] This data is presented in graphic form in Appendix F (Figures F.1-F.3)

similar to the changes in probability of *no concern,* and the probability of *partial concern* was also similar.

This pattern of findings was also found in parent reported RRBIs, but not across the other sources. Specifically, for diagnostician observed RRBIs, the LOR for *no concern* (versus *concern/partial concern*) was substantially smaller than the LOR for *concern* (versus *no/partial concern*). This indicates that the increase in the probability of *concern* was not only due to a decrease in the probability of *no concern*, but also a decrease in *partial concern*. A negligible difference in the probability of *no concern* in diagnostician observed RRBIs could not be excluded with 80% confidence. Even though $P_{(meaningful)}$ supports a non-trivial decrease in *no concern* between assessments, examination of the LOR itself shows that the size of the effect for *concern* was roughly 2.8 times that for *no concern*. There was no strong evidence of a difference in the probability that *concern* was raised (or *no concern* was raised) in the teacher-report RRBI data (neither a negligible difference nor a small difference in the opposite direction could be excluded with 80% confidence in either case). However, the $P_{(meaningful)}$ probabilities suggest some evidence of a non-trivial change (both in *concern* and *no concern*). The similar LORs across *concern* and *no concern* suggest that the increased probability of *concern* was mirrored in the decreased probability of *no concern*, and that the probability of partial concern was similar across assessments. The HDIs for teacher report are generally wider than for other report sources, likely because fewer variables were considered from teacher report than the other sources, leading to considerable uncertainty in these estimates.

Criteria Met

Table 4.3 presents the results of the conditional logistic regressions, where the log odds ratios refer to the outcome of a criterion being met versus partially/not met. Children were more likely to meet each of the criteria at Assessment 2 (i.e., the assessment at which

they received an ASD diagnosis). There was no conclusive evidence that the child's sex/gender alone was associated with greater probability that they would meet any criterion at either assessment. However, there was weak evidence of interactions between assessment number and sex/gender in whether or not Criteria B2 and B3 were met. There was a large increase in the proportion of females who met Criterion B2 across the assessments (81.0%) and this was larger than the increase in the proportion of males (66.7%). Indeed, for females,

Table 4.3

Logistic Regression Effect of Assessment Number, Sex/Gender and Their Interaction on Criteria Met/Not Met

	Effect of Assessment No.	Effect of Sex/Gender		Interaction	
Criterion	LOR [$HDI_{80\%}$] [a]	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Domain A	8.51 [5.42, 11.37]	0.11 [-2.21, 2.50]	.50	0.41 [-4.11, 4.67]	.54
A1	8.86 [3.01, 15.30]	0.17 [-0.94, 1.61]	.54	-0.06 [-2.22, 1.93]	-.48
A2	10.27 [3.88, 19.09]	-0.04 [-1.49, 1.25]	-.46	0.10 [-1.76, 2.57]	.50
A3	6.98 [2.73, 11.46]	0.15 [-0.86, 1.46]	.53	-0.04 [-2.00, 1.74]	-.47
Domain B	8.38 [5.13, 11.40]	-0.58 [-3.00, 1.84]	-.60	1.11 [-3.23, 5.72]	.62
B1	2.53 [1.13, 3.86]	-0.36 [-1.48, 0.49]	-.65	0.68 [-0.57, 2.40]	.71
B2	3.93 [2.52, 5.32]	0.19 [-0.45, 1.17]	.56	-1.01 [-3.07, 0.48]	-.77
B3	2.71 [1.36, 3.98]	0.41 [-0.40, 1.53]	.67	0.93 [-0.47, 2.75]	.76
B4	2.45 [1.12, 3.69]	0.11 [-0.63, 1.08]	.50	0.39 [-0.66, 1.98]	.63

Note. Criterion A1: deficits in social-emotional reciprocity; A2: deficits in nonverbal communication behaviours; A3: deficits in developing, maintaining, and understanding relationships; B1: stereotyped/ repetitive motor movements, use of objects, or speech; B2: insistence on sameness, routines, or ritualised behaviour; B3: restricted and fixated interests; B4: hyper-/hypo-reactivity to sensory input. Positive LORs (assessment no.) = greater probability of being met at Assessment 2, Positive LORs (sex/gender) = greater probability of being met for males (these directions are identical for $P_{(meaningful)}$). LOR = log odds ratio. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction.

[a] For each criterion, there was a 1.00 probability of the LOR posterior falling outside the ROPE.

Criterion B2 was the least likely RRBI criterion to be met at Assessment 1 (probability = 0.10, $HDI_{80\%}$ = [0.00, 0.19]), but the most likely to be met at Assessment 2 (probability = 0.91, $HDI_{80\%}$ = [0.83, 1.00]). A similar proportion of males and females met Criterion B3 at Assessment 1 (13.3% and 13.7%, respectively), but the proportion increase by Assessment 2 was 23.1% higher for males (66.7%) than females (43.6%). This suggests that for B2, females were more likely to have difficulties emerge between assessments, but for B3, males were more likely to have difficulties emerge. For the remaining criteria, the evidence did not allow for any conclusions (even weak) regarding interactions between assessment number and sex/gender.

The proportions of children meeting and showing *any* difficulty in each criterion (i.e., criterion met or partially met) were calculated from the logistic regression models and are presented in Table 4.4. Relative risks are also included to aid in interpretation. As depicted below, Criterion B3 was the area in which both males and females were most likely to increase in probability of demonstrating *any* atypicality (i.e., criterion met or partially met), but this increase was larger for females. The greatest increase in probability of developing clinically significant atypicality across assessments was in Criterion B2 for females (81.0% higher, relative risk = 9.02) and Criterion A2 for males (88.7%, relative risk = 8.85). Figure 4.1 visually represents the proportions of children who (a) met, (b) partially met, or (c) did not meet, each criterion at each assessment.

Table 4.4

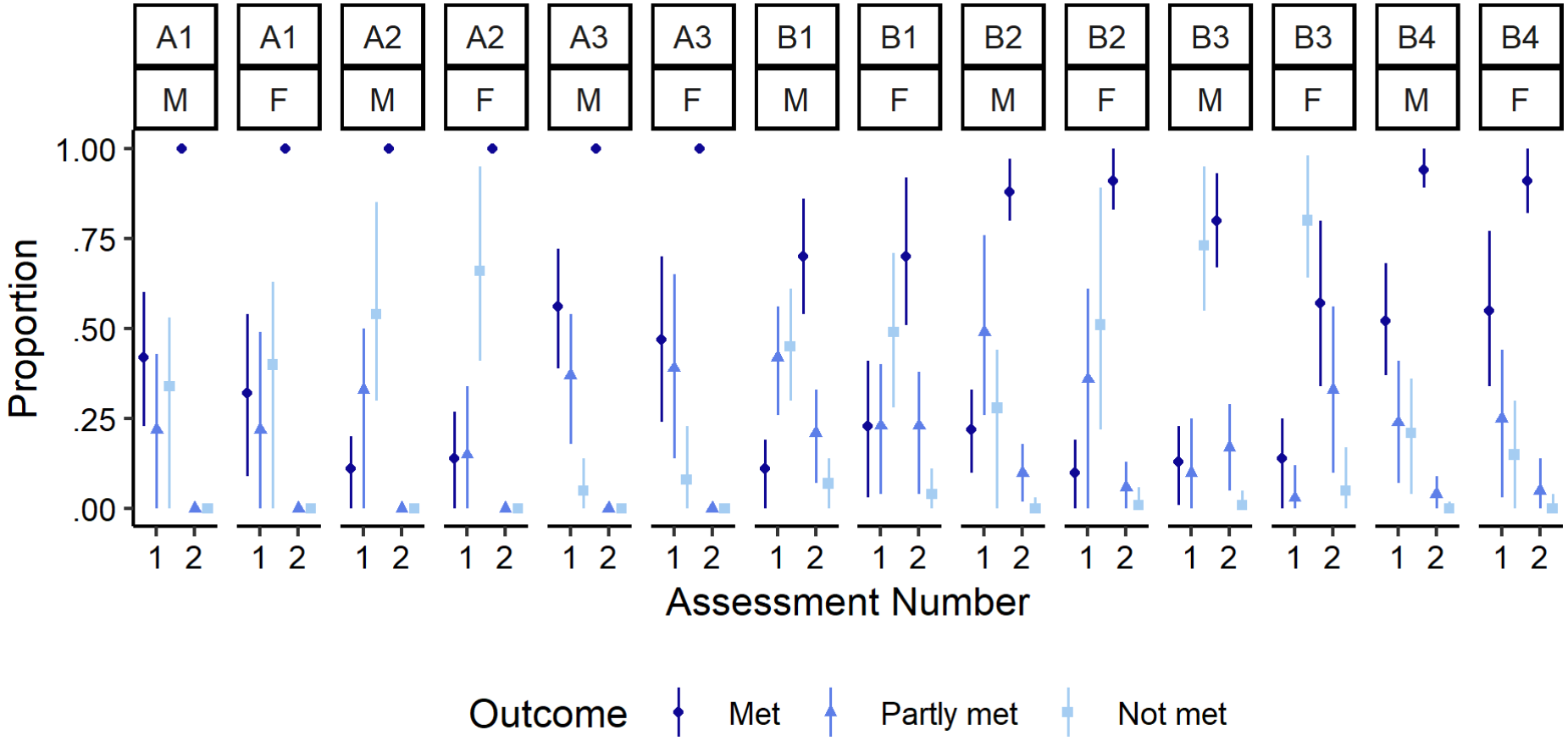
Descriptive Statistics and Relative Risk for Criteria Met/Partially Met at Assessment 1 and 2

Variable		Criterion						
		A1	A2	A3	B1	B2	B3	B4
Males								
Proportion meeting criterion (%)	Ax. 1	41.5	11.3	55.5	10.7	21.5	13.3	52.0
	Ax. 2	100.0	100.0	100.0	69.8	88.0	80.0	94.3
	Relative risk	2.41	**8.85**	1.80	6.52	4.09	6.02	1.81
Proportion with any atypicality (%)	Ax. 1	63.0	43.9	92.0	52.6	70.5	23.5	76.3
	Ax. 2	100.0	100.0	100.0	90.7	98.2	96.9	98.7
	Relative risk	1.59	2.28	1.09	1.72	1.39	**4.12**	1.29
Females								
Proportion meeting criterion (%)	Ax. 1	32.4	14.0	47.1	23.2	10.1	13.7	54.9
	Ax. 2	100.0	100.0	100.0	70.3	91.1	57.3	91.1
	Relative risk	3.09	7.14	2.12	3.03	**9.02**	4.18	1.66
Proportion with any atypicality (%)	Ax. 1	54.5	28.9	86.2	46.6	46.4	16.4	80.1
	Ax. 2	100.0	100.0	100.0	93.1	96.6	89.8	97.4
	Relative risk	1.83	3.46	1.16	2.00	2.08	**5.48**	1.22

Note. Proportion with any atypicality indicates that the criterion was met or partially met. All proportions were derived from the logistic regression models. Relative risk was calculated as follows: Proportion at Assessment 2 ÷ proportion at Assessment 1. The largest relative risk in each set of results is highlighted in boldface. Ax. = Assessment.

Figure 4.1

Proportion of Criteria Met, Partially Met and Not Met by Sex/Gender and Assessment Number



Note. Error bars represent HDIs $_{(80\%)}$. M = male, F = female.

Social Communication

Parent Report. Results of logistic regressions of parent-report behaviours for their daughters are presented in Table 4.5 (assessment number = repeated predictor variable, behaviour = outcome variable). Across a large number of areas, there was a meaningfully higher probability that social difficulty would be reported by parents of females at Assessment 2 compared to Assessment 1. In Criterion A1 (deficits in social/emotional reciprocity), females had a 2.3 times higher probability of demonstrating atypical *Social approach* (i.e., initiation and response to social interaction) and a 3.6 times higher probability of demonstrating atypical *Conversation content* (i.e., tangential statements, monologuing behaviours). In Criterion A2 (deficits in nonverbal communication), the largest increase was seen in the probability that difficulties with *Eye contact* would be reported (11.0 times higher probability at Assessment 2). Similarly, abnormality in *Facial expression* (e.g., exaggerated or flat affect), was 24.0 times more likely to be reported at Assessment 2 than Assessment 1. However, meaningful increases in Criterion A3 features were most numerous. There was strong evidence of an increased probability of reporting difficulty in four of the seven areas analysed. The greatest increase was in *Imagination or spontaneity in play* (8.2 times higher at Assessment 2), followed by *Possessiveness of objects/difficulty losing in games* (5.6 times higher), *Friendship maintenance* (1.4 times higher), and *Submissiveness or domination in play* (1.2 times higher). Atypicality in *Conversation content* is the only area within the Social Communication domain in which a negligibly sized difference in probability of reporting difficulty between assessments can be excluded with 80% confidence.

Table 4.5

Logistic Regression Effect of Assessment Number on Social Communication Difficulties: Parent Report

Behavioural Category	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Social approach	2.10 [-0.41, 5.03]	.90
Adherence to social norms	0.84 [-0.30, 2.08]	.79
Reciprocal conversation	0.34 [-0.47, 1.38]	.64
Sharing interests	0.61 [-0.48, 2.01]	.71
Sharing emotions	0.20 [-0.68, 1.45]	.56
Content of conversation	**2.20 [0.57, 3.65]**	**.98**
Literal language	0.09 [-0.77, 1.17]	.49
Eye contact	3.36 [-0.11, 6.88]	.98
Use of nonverbal communication	0.00 [-3.15, 3.38]	.45
Facial expression	3.48 [-0.05, 6.85]	.97
Nonverbal understanding	0.87 [-0.28, 2.11]	.80
Response to nonverbal behaviour	0.85 [-0.27, 2.17]	.79
Emotional regulation	0.00 [-0.87, 0.94]	.40
Imagination/spontaneity in play	2.49 [-0.16, 5.02]	.95
Submissive/dominating in play	1.71 [-0.37, 3.85]	.88
Possessiveness/difficulty losing	2.15 [-0.50, 5.49]	.90
Friendship formation	1.14 [-0.67, 3.37]	.79
Friendship maintenance	1.53 [-0.10, 2.87]	.91
Social interest	1.05 [-0.26, 2.46]	.82
Consistent companionship	0.37 [-0.44, 1.45]	.65

Note. Positive LORs = greater probability of being reported at Assessment 2, negative LORs = greater probability at Assessment 1. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

Diagnostic Observations. There was a meaningful increase in the probability that a female would demonstrate observable atypicality at Assessment 2 compared to Assessment 1 in approximately half of the areas analysed (see Table 4.6). For two particular areas, namely *Content of conversation* and *Imaginative play,* both a difference of zero and a negligible difference in the probability of demonstrating significant atypicality can be excluded with 80% confidence. No females were deemed to demonstrate significant atypicality in *Conversation content* at Assessment 1, but for four females, this was considered somewhat atypical. In contrast, at Assessment 2, half of the females demonstrated significant difficulty. Similarly, two females showed partial atypicality in *Imaginative play* at Assessment 1, whereas four showed significant atypicality and one partial atypicality in this area at Assessment 2.

Table 4.6

Logistic Regression Effect of Assessment Number on Social Communication Difficulties: Diagnostic Observations

Behavioural Category	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Social approach	1.14 [-0.13, 2.27]	.87
Adherence to social norms	1.14 [-0.13, 2.28]	.87
Reciprocal conversation	3.16 [-0.11, 6.58]	.98
Sharing interests	0.60 [-0.49, 1.98]	.71
Sharing emotions	0.78 [-1.17, 3.63]	.69
Content of conversation	**4.87 [0.96, 9.11]**	**.99**
Literal language	2.85 [-0.42, 6.74]	.92
Eye contact	0.81 [-0.30, 2.06]	.79
Use of nonverbal communication	1.73 [-0.61, 4.79]	.84
Facial expression	-0.00 [-1.01, 0.95]	-.41
Nonverbal interpretation	1.74 [-0.02, 2.95]	.94
Understanding of friendship	3.34 [-0.04, 6.81]	.98
Inclusive of assessor	0.25 [-0.74, 1.60]	.58
Imaginative play	**4.20 [0.41, 7.95]**	**.99**

Note. Positive log odds ratios = greater probability of being reported at Assessment 2, negative ratios = greater probability at Assessment 1. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

Teacher Report. Teacher-report data were available for eight females at their first assessment and second assessment. The change in the probability that significant concern was raised in each area across assessments is presented in Table 4.7. The only area in which there was evidence of a large increase in the probability of teacher concern was *Conversation skills*, but neither a negligible difference, nor equivalent concern, can be excluded with 80% confidence.

Table 4.7

Logistic Regression Effect of Assessment Number on Social Communication Difficulties: Teacher Report

Behavioural Category	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Social approach	0.40 [-0.76, 2.21]	.63
Conversation skills	3.84 [-0.23, 8.08]	.96
Nonverbal interpretation	0.50 [-0.54, 1.91]	.68
Use of nonverbal communication	-0.06 [-1.19, 0.90]	-.47
Friendship formation	-1.01 [-3.53, 0.78]	-.76
Friendship maintenance	0.00 [-0.93, 0.94]	.41

Note. Positive LORs = greater probability of being reported at Assessment 2, negative LORs = greater probability at Assessment 1. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

Repetitive and Restricted Behaviours and Interests

Parent Report. Analysis of RRBI data (Table 4.8) revealed a 4.6 times higher probability of parents reporting *Stereotypical movement* at Assessment 2, which was greater than the increase for both *Stereotypical speech* and *Stereotypical object use* (both had a 1.8 times higher probability at Assessment 2). However, for all areas, an effect in either direction could not be excluded with 80% confidence. There were large increases in the probability of reporting *Difficulty managing change* (11.5 times higher), difficulty in *Routine adherence* (6.3 times higher) and a smaller increase in the probability of reporting *Rigid thinking* (1.85 times higher) in Criterion B2. Finally, there was a 12.0 times increase in the probability of *Oral avoiding* behaviours, such as aversion to certain tastes, between assessments. In contrast, the probability of *Auditory seeking* behaviours decreased by a factor of 6.0 across assessments.

Table 4.8

Logistic Regression Effect of Assessment Number on RRBI Behaviours: Parent Report

Behavioural Category	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Stereotypical movement	2.05 [-0.44, 4.98]	.89
Stereotypical speech	0.84 [-0.28, 2.13]	.79
Stereotypical object use	0.82 [-0.30, 2.06]	.79
Difficulties with change	**4.85 [1.26, 9.01]**	**1.00**
Routine adherence	**2.89 [1.21, 4.49]**	**.99**
Task switching	0.50 [-0.48, 1.72]	.69
Rigid thinking	3.03 [-0.03, 5.53]	.98
Auditory seeking	-1.80 [-4.96, 0.64]	-.84
Auditory avoiding	1.25 [-0.64, 3.88]	.80
Tactile seeking	0.50 [-0.52, 1.74]	.69
Tactile avoiding	-0.00 [-0.91, 0.89]	-.40
Olfactory seeking	-0.79 [-3.60, 1.19]	-.70
Olfactory avoiding	0.00 [-1.84, 1.81]	-.43
Oral seeking	-0.49 [-1.75, 0.47]	-.69
Oral avoiding	3.17 [-0.03, 6.11]	.98
Visual seeking	0.00 [-1.26, 1.19]	-.42
Visual avoiding	0.80 [-1.25, 3.66]	.70

Note. Positive LORs = greater probability of being reported at Assessment 2, negative LORs = greater probability at Assessment 1. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

There was evidence to suggest the emergence of a large number of specific stereotypical behaviours between Assessments 1 and 2 (i.e., a higher probability of being reported at the second assessment; Table 4.9). In particular, *Self-injurious behaviour* and the production of *Unusual noises* were each six times more likely to be present at Assessment 2. For a small number of behaviours (i.e., *Toe walking*, *Hand flapping,* and *Echolalia*), the

Table 4.9

Logistic Regression Effect of Assessment Number on Stereotypical Behaviours (Criterion B1): Parent Report

Behavioural Category	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Motor		
Toe walking	-1.10 [-3.09, 0.95]	-.75
Flapping	-2.58 [-5.87, 0.77]	-.86
Spinning	2.62 [-0.68, 6.00]	.86
Rocking/jumping	**4.68 [1.28, 8.06]**	**.99**
Physical rigidity	0.01 [-2.26, 2.41]	.48
Hand mannerisms	**4.66 [0.89, 8.38]**	**.97**
Self-injurious	**5.66 [2.01, 8.94]**	**.99**
Speech/language		
Echolalia	-0.98 [-2.87, 0.90]	-.74
Neologisms	-0.00 [-1.48, 1.54]	-.47
Pronoun reversal	**4.62 [0.76, 8.19]**	**.97**
Repetitive speech	**2.23 [0.55, 3.86]**	**.97**
Accents	1.12 [-0.86, 3.14]	.76
Unusual noises	**5.55 [2.01, 8.92]**	**.99**
Odd prosody	-0.01 [-2.20, 2.09]	-.48
Object use		
Lining up	**4.67 [1.07, 7.96]**	**.99**
Grouping	**1.92 [0.24, 3.63]**	**.94**
Spinning/flicking/pushing	**4.60 [1.17, 7.94]**	**.99**
Repetitive play	0.00 [-1.50, 1.54]	.47
Deconstruction	3.45 [-0.43, 7.34]	.89

Note. Positive LORs = greater probability of being reported at Assessment 2, negative LORs = greater probability at Assessment 1. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

probability of being reported was higher at Assessment 1, indicating that some females had ceased these behaviours before Assessment 2. There was no evidence of a difference in the probability that *Physical rigidity*, use of *Neologisms*, *Odd prosody* of speech or *Repetitive play* were reported at either assessment.

No specific stereotypical behaviours were reported by parents of only two females (16.7%) at the first assessment, with parents reporting an average of between three and four stereotypies (*SD* = 2.68). In contrast, all females, regardless of whether they met the B1 criterion overall, had a minimum of three specific stereotypical behaviours reported at Assessment 2 and approximately six stereotypical behaviours on average (*SD* = 1.90). Slightly over twice the number of motor mannerisms and 1.7 and 1.9 times as many speech and object use mannerisms (respectively) were reported by parents at reassessment compared with the initial presentation.

At least one restricted interest was reported for 75% of females at their first presentation, with random objects such as shells, rocks, or feathers emerging as the most common. In contrast, restricted interests were reported for all females at Assessment 2, with *Specific fictional characters or programs* being the most common and reported for 58.3%. Restricted interests in five areas were more commonly reported at Assessment 2. The strongest evidence of increase was for restricted interest in *Screens* such as television, computer, and tablet devices (LOR = 6.00, $HDI_{80\%}$ = [2.27, 9.57], $P_{(meaningful)}$ = 100%). Similarly, females were more likely to present with restricted interests in *Specific people* (LOR = 4.74, $HDI_{80\%}$ = [1.02, 8.52], $P_{(meaningful)}$ = 97.5%), *Animals* (LOR = 3.95, $HDI_{80\%}$ = [0.47, 7.21], $P_{(meaningful)}$ = 96.4%), *Craft* (LOR = 3.80, $HDI_{80\%}$ = [0.43, 7.15], $P_{(meaningful)}$ = 96.1%), a *Specific program or character* (LOR = 1.84, $HDI_{80\%}$ = [0.18, 3.52], $P_{(meaningful)}$ = 94.3%), and *Particular toys* (LOR = 1.49, $HDI_{80\%}$ = [-0.24, 3.22], $P_{(meaningful)}$ = 87.1%) at Assessment 2. On the other hand, there was an approximately equivalent probability that a

parent would report restricted interests in seemingly *Random objects* or *vehicles* (including toy vehicles) across assessments. Overall, females generally presented with both a larger number and a broader range of restricted interests at Assessment 2.

Diagnostic Observations. The results of logistic regressions for diagnostic observations are presented in Table 4.10. Stereotypical and unusual sensory behaviours were approximately as likely to be observed at either assessment. However, there was relatively strong evidence of an increase in the probability that a clinician would observe difficulty in *Transitioning* between tasks and evidence of *Restricted interests* at Assessment 2 compared to Assessment 1. There was moderate evidence of this for *Rigid thinking* and the presence of *Routines and rituals* during assessment. However, neither a negligible difference nor a difference of zero can be excluded with 80% confidence based on these results.

Table 4.10

Logistic Regression Effect of Assessment Number on RRBI Behaviours: Diagnostic Observations

Behavioural Category	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Stereotypical movement	0.00 [-0.99, 0.95]	.41
Stereotypical speech	0.24 [-0.77, 1.61]	.58
Stereotypical object use	-0.01 [-3.27, 3.48]	-.45
Transitioning	2.91 [-0.45, 6.90]	.93
Rigid thinking	1.78 [-0.38, 4.00]	.88
Routines and rituals	1.93 [-0.70, 5.30]	.85
Restricted interests	3.09 [-0.60, 8.13]	.93
Sensory behaviours	0.17 [-0.65, 1.29]	.55

Note. Positive LORs = greater probability of being reported at Assessment 2, negative LORs = greater probability at Assessment 1. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

Teacher Report. As shown in Table 4.11, no conclusive evidence was found to suggest that the probability of teacher concern in any RRBI behaviour examined differed between Assessment 1 and 2.

Table 4.11

Logistic Regression Effect of Assessment Number on RRBI Behaviours: Teacher Report

Behavioural Category	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Stereotypical movement	0.08 [-0.98, 1.37]	.49
Stereotypical speech	-0.01 [-3.18, 3.09]	-.45
Stereotypical object use	0.73 [-1.23, 3.64]	.68
Routines and rituals	0.66 [-1.26, 3.40]	.67
Difficulties with change	0.22 [-0.59, 1.37]	.57
Restricted interests	0.68 [-1.25, 3.49]	.67
Sensory behaviours	-0.79 [-3.32, 1.06]	-.70

Note. Positive LORs = greater probability of being reported at Assessment 2, negative LORs = greater probability at Assessment 1. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

Interim Discussion

Study 2a examined changes in the presentation of females who undertook two formal developmental assessments and received an ASD diagnosis only at the second presentation. Reasons why an ASD diagnosis was not made at the first assessment, and concerns leading to reassessment, were also explored. Finally, the pattern of met and unmet ASD criteria was compared between males and females who received an ASD diagnosis, but only after an initial negative result.

The most common reason that both males and females were not provided with ASD diagnoses at Assessment 1 was that they did not demonstrate enough ASD-related atypicality.

However, the specific areas in which females lacked sufficient atypicality generally differed from those of males. Specifically, females were considered to be more socially able than their male counterparts. This is consistent with the finding that for females, many specific social difficulties emerged according to parent report and diagnostician observation that had not been present at the initial assessment. In particular, significant difficulties with *Conversation content* were meaningfully likely to be reported at the second assessment than the first, across all report sources. These findings support the hypothesis that females' social difficulties may either (a) emerge, or (b) become more salient as social demands increase beyond their abilities in later childhood (Hsiao et al., 2013; Mandy et al., 2018).

Males and females also differed in the parent-reported concerns cited as reasons for reassessment. It is surprising that increased anxiety was more likely to be a concern raised for males, as internalising difficulties have frequently been found to be higher among autistic females (Solomon et al., 2012). Indeed, the only area of concern that was more commonly cited for females was *General concern.* This reflects increased concern across diverse areas of functioning and was mirrored in the increased probability of reporting many specific difficulties at Assessment 2.

The increase in probability of demonstrating clinically significant difficulty differed by sex/gender and criterion. For females, the largest increase in probability of meeting a criterion contributing to diagnostic outcome was resistance to change, routines, and rituals (Criterion B2). Indeed, of all restricted and repetitive behaviours, females were least likely to meet Criterion B2 at Assessment 1, but most likely to meet it at Assessment 2. The emergence of females' difficulties in resistance to change was also reflected in particular behaviours. The reason for the emergence of these difficulties is unknown. However, others have hypothesised that this may be the result of strengthening neuro-cognitive processes underpinning resistance to change (Gomot & Wicker, 2012) or increase in moderating

difficulties, such as anxiety (which has been associated with insistence on sameness; Uljarevic et al., 2017). Alternatively, it is possible that these difficulties may simply become more ingrained, salient, or functionally disruptive with time. It should also be noted that increased parent-reported concern may reflect changes in parents' understanding or awareness of symptoms, and therefore the extent to which they are reported at reassessment.

Several specific concerns were more likely to be reported across all sources for females at Assessment 2, but others were found to be almost as likely at Assessment 1 (e.g., an absence of consistent companions). Further research should investigate these trends with a larger sample in order to clarify whether less severe difficulty in criteria such as B2 among young girls with other ASD traits is sufficient for diagnosis (i.e., lowering the diagnostic threshold for younger girls), with the expectation that difficulties may increase. Alternatively, it may be beneficial for clinicians to conduct follow-up assessments for females who present with less severe difficulties in conversation content, or resistance to change, but with other ASD traits, with the expectation that these latter behaviours may emerge. While it is critical that early 'red flags' are not overlooked, care must be taken not to prematurely diagnose this lifelong developmental condition and pathologise those who are managing their characteristics. Sex/gender differences in the developmental trajectories of specific ASD-related behaviours are not yet well understood, and therefore it will be important for future research to clarify predictors of increased difficulty in these areas over time.

Study 2b

Method

Participants

Data were extracted from the ASD diagnostic assessment reports of 222 children (n = 98 females and n = 124 males) who had undertaken an ASD assessment and either (a) received an ASD diagnosis (n = 156; females: 50.0%, n = 78; males: 50.0%, n = 78, hereafter referred to as the ASD group), or (b) did not receive an ASD diagnosis, despite the presence of ASD traits (n = 66; females: 30.3%, n = 20; males: 69.7%, n = 46, hereafter referred to as the non-ASD group). All participants were clients of the same private clinic as Study 2a and were assessed according to the protocol outlined in Study 2a.[38] Participants were excluded if their cognitive ability was in the range of intellectual disability (≤70).[39]

A total of 79 children were identified as having received a negative result at an ASD assessment. In order to restrict the sample to children with many ASD traits but who did not receive a diagnosis (hence *narrowly below* the diagnostic threshold), the following inclusion criteria were applied:

1. At assessment, no alternative diagnostic explanation (e.g., reactive attachment disorder, post-traumatic stress disorder, intellectual disability) was given to *entirely* explain the presenting difficulties (n = 10 excluded);

[38] The participants of Study 2a (i.e., children diagnosed with ASD at a second assessment) were regarded as special cases and so were considered separately. Data from the initial presentations (negative to ASD) of children later diagnosed have not been included in this study. It is acknowledged that some children in the non-ASD groups of Study 2b may later receive ASD diagnoses.

[39] An attempt was made to statistically control for intellectual ability in the analysis. However, this was not practical given that IQ data were only available for n = 143 participants (64.4%). Children with intellectual disabilities were, therefore, excluded.

2. At assessment, the child was deemed to at least partially meet one social communication criterion and one RRBI criterion (n = 3 excluded); and
3. The child had not received a diagnosis of ASD since the time of their assessment and data collection.

Procedure

The research procedure for Study 2b was identical to that implemented in Study 2a. However, fine-grained data were collected for both males and females in each condition in order to address the more specific research questions. Parent-report and diagnostician observation data were available for all children, but teacher-report data were only available for 71.2% of children (n = 158; females: 64.3%, n = 63; males: 76.6%, n = 95). The CARS2 (ST and HF) and GARS-3 were the most commonly used standardised instruments and therefore participants' scores on these instruments were extracted if available. Following data collection, an independent rater, blind to the study aims, re-coded 10% of the data and there was between 78.3% and 100% agreement across variables (M = 86.8%). Intra-rater reliability checks showed between 87% and 100% consistency across variables.[40]

Data Analysis

Bayesian logistic regression analysis was again applied, and log odd ratios used as the basis for inference, with relative risk reported to aid interpretation. The assessment decision (i.e., positive or negative ASD result) and the child's sex/gender were defined as predictor variables (between groups). The presence of any given ASD behaviour was coded in binary form: (0) *absent/no concern raised* or (1) *present/concern raised*.[41] Age (centred and scaled to SD = 1) was included as a covariate.

---

[40] For a description of the operationalisation of each behaviour, refer to Appendix E.

[41] Initially, the outcome variables (ASD behaviours) were coded in tertiary form i.e., (0) *no concern*, (1) *partial/some concern* or (2) *concern*. However, due to reliability concerns and the relatively

The interaction between assessment result and sex/gender on whether any particular behaviour was reported was of particular interest in the present study. In referring to interactions, relative risks are reported. However, in some instances, the relative risk denominators differed considerably between males and females, rendering them difficult to compare. In these cases, it is more valuable to consider proportion differences reported in the tables (i.e., the proportion of ASD males for whom the behaviour was reported minus the proportion of non-ASD males for whom the behaviour was reported; and likewise for females).

Simple main effects were also examined in order to identify whether ASD or non-ASD children contributed more towards any given sex/gender difference (i.e., probability for ASD males minus females; probability for non-ASD males minus females). Simple main effects are displayed in the tables and in the text, where noteworthy.

Age at Assessments

A Bayesian ANOVA revealed meaningful differences in the ages of children in each group (refer to Table 4.12 for descriptive information on participants' age and cognitive ability). Of children diagnosed with ASD, females were meaningfully older (d = -0.67, $HDI_{80\%}$ = [-0.87, -0.47], $P_{(meaningful)}$ = 100%). However, there was equivocal evidence of a difference in age between non-ASD children (d = -0.10, $HDI_{80\%}$ = [-0.45, 0.23], $P_{(meaningful)}$ = 50.3%). Males in the non-ASD group were older than diagnosed males (d = 0.32, $HDI_{80\%}$ = [0.08, 0.53] $P_{(meaningful)}$ = 89.7%). There was weak evidence that non-ASD females were older than diagnosed females (d = 0.25, $HDI_{80\%}$ = [-0.04, 0.53], $P_{(meaningful)}$ = 75.4%).

---

restricted sample size, these were collapsed to binary variables, with (1) *partial concern* merged with (0) *no concern*.

Table. 4.12

Age and Intellectual Functioning of Children by Sex/Gender and Assessment Result

Group	Variable	Males	Females	All
ASD children	*n*	78	78	156
	Age (years)			
	M (SD)	6.90 (3.26)	9.51 (4.32)	8.19 (4.06)
	Range	2.08-16.58	1.83-17.83	1.83-17.83
	IQ			
	M (SD)	96.88 (12.11)	96.67 (15.23)	96.94 (14.15)
	Range	76-136	70-131	70-136
Non-ASD children	*n*	46	20	66
	Age (years)			
	M (SD)	8.15 (2.96)	8.53 (4.53)	8.31 (3.48)
	Range	2.00-14.16	1.50-17.83	1.50-17.83
	IQ			
	M (SD)	95.38 (15.19)	97.00 (13.16)	95.73 (14.39)
	Range	70-137	78-126	70-137
All	*n*	124	98	222
	Age (years)			
	M (SD)	7.33 (3.21)	8.31 (3.48)	8.23 (3.89)
	Range	2.00-16.58	1.50-17.83	1.50-17.83
	IQ			
	M (SD)	96.07 (13.52)	95.73 (14.39)	96.47 (14.19)
	Range	70-137	70-131	70-137

Cognitive Ability

Standardised information regarding cognitive ability was available for 70.4% of females and 59.7% of males (60.9% of ASD children and 72.7% of non-ASD children). This information had been collected at the time of (or prior to) diagnostic assessment, using either the Wechsler Intelligence Scale for Children (Wechsler, 2003; 2014) or Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 2002; 2012). In the event that only certain subtest or index scores were available, Full Scale Intelligence Quotients (FSIQs) were prorated (19.6% of cases). A Bayesian ANOVA on the subsample for which data were

available revealed no difference in FSIQ between males and females with ASD (d = 0.01, $HDI_{80\%}$ = [-0.22, 0.25], $P_{(within\ ROPE)}$ = 43%), but there was equivocal evidence that non-ASD males had lower cognitive ability than non-ASD females (d = -0.10, $HDI_{80\%}$ = [-0.41, 0.19], $P_{(meaningful)}$ = 50%). There was no conclusive evidence that ASD females differed from non-ASD females in cognitive ability (d = -0.02, $HDI_{80\%}$ = [-0.32, 0.26], $P_{(meaningful)}$ = 36.1%). Similarly, there was no evidence that non-ASD males and ASD males differed in cognitive ability (d = 0.10, $HDI_{80\%}$ = [-0.15, 0.37], $P_{(meaningful)}$ = 50.3%).

Family History

Information regarding family history of ASD was available for 85.0% of children. There was strong evidence that children who received ASD diagnoses had a higher probability of having a family history of ASD (LOR = 0.70, $HDI_{80\%}$ = [0.27, 1.13], $P_{(meaningful)}$ = 96.4%). In general, females had a higher probability of having a family member with ASD (LOR = -0.85, $HDI_{80\%}$ = [-1.28, 0.41], $P_{(meaningful)}$ = 98.8%), but a positive family history was more strongly associated with receiving an ASD diagnosis for males (LOR = 0.76, $HDI_{80\%}$ = [-0.13, 1.59], $P_{(meaningful)}$ = 83.5%).

Referral Pathway

The relationship of the referrer to the child was categorised as family or a professional (e.g., medical doctor, psychologist, speech pathologist, or the school). A child had a higher probability of being referred by a professional if they were ultimately diagnosed with ASD (LOR = 1.02, $HDI_{80\%}$ = [0.48, 1.57], $P_{(meaningful)}$ = 98.4%), and if they were male (LOR = 0.57, $HDI_{80\%}$ = [0.03, 1.11], $P_{(meaningful)}$ = 86.9%). Here, a negligible effect of sex/gender cannot be excluded with 80% confidence.

Previous Psychiatric Diagnoses and Allied Health Support

Females had a higher probability of having received a previous psychiatric diagnosis than males (LOR = -0.63, $HDI_{80\%}$ = [-1.20, 0.08], $P_{(meaningful)}$ = 88.7%; a negligibly sized

difference cannot be excluded with 80% confidence). Similarly, children subsequently diagnosed with ASD had a meaningfully higher probability of having received a previous diagnosis than children who were not diagnosed with ASD (LOR = 0.73, $HDI_{80\%}$ = [0.18, 1.31], $P_{(meaningful)}$ = 93.0%).

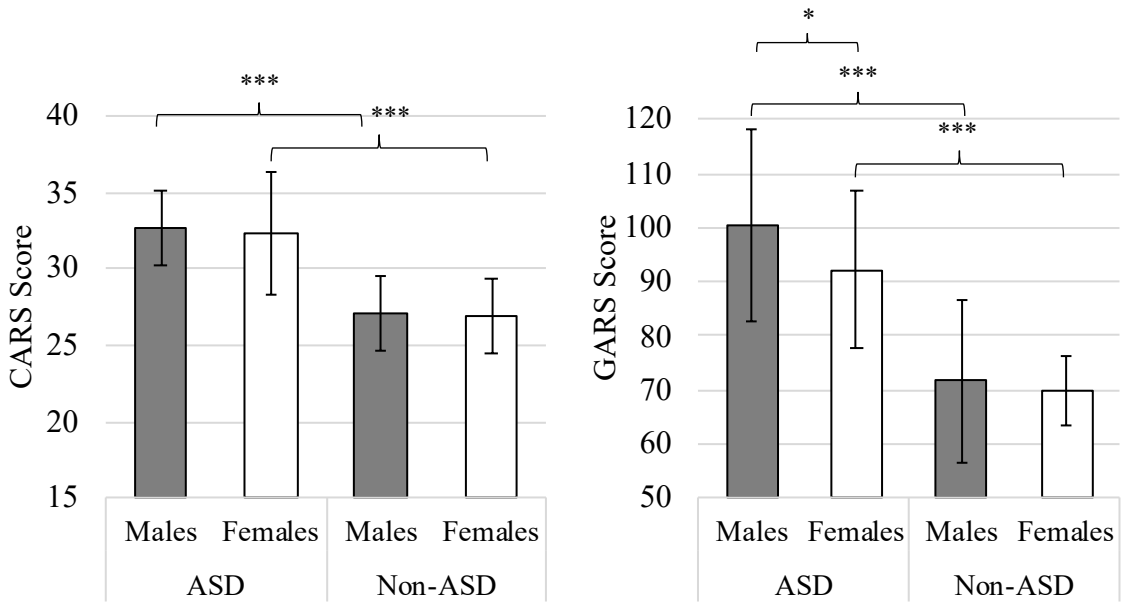
Whether a child had previously accessed psychology, speech pathology, or occupational therapy support was coded using a binary variable (*Yes* or *No*) and this information was available for 92.3% of children (*n* = 205; 88 females). Sex/gender, but not assessment result, meaningfully predicted the probability of a child having had previous psychology support (females with a higher probability: LOR = -0.51, $HDI_{80\%}$ = [-.92, 0.10], $P_{(meaningful)}$ = 89.1%). Males without previous psychology input had a higher probability of receiving ASD diagnosis (17.3%), but females were slightly less likely to have accessed psychology if they did not receive an ASD diagnosis (4.2%). While ASD assessment result did not predict whether a child had previously received speech pathology input, previous speech pathology was strongly associated with male sex/gender (LOR = 0.85, HDI = [0.40, 1.29], $P_{(meaningful)}$ = 98.7%). Autistic males had the highest probability of having had occupational therapy support (37.0%), followed by non-ASD females (29.1%) and then autistic females and non-ASD males (19.0% each).

Results

CARS2 (ST or HF) scores were available for 60.9% of ASD children (*n* = 95; *n* females = 46) and 34.8% of non-ASD children (*n* = 23; *n* females = 7). A Bayesian ANOVA revealed meaningful differences in scores across the four groups (Figure 4.2, left panel).

Figure 4.2

Mean Scores on CARS2 and GARS-3 by Sex/Gender and Assessment Result



Note. Error bars indicate standard deviations. * small-medium effect (*d* between 0.2 and 0.5), ** medium effect (0.5 to 0.8), *** strong effect (> 0.8; Cohen, 1988).

As expected, there was a large difference in CARS2 score between females who did (M = 32.35, SD = 3.96) and did not receive ASD diagnoses (M = 26.85, SD = 2.47; d = 1.44, $HDI_{80\%}$ = [1.08, 1.79], $P_{(meaningful)}$ = 100%). The same pattern was found between autistic (M = 32.61, SD = 2.45) and non-ASD males (M = 27.11, SD = 2.50; d = 1.44, $HDI_{80\%}$ = [1.16, 1.72], $P_{(meaningful)}$ = 100%). No meaningful differences were found between males and females with ASD (d = 0.07, $HDI_{80\%}$ = [-0.13,0.27], $P_{(meaningful)}$ = 41%) nor non-ASD males and females (d = 0.06, $HDI_{80\%}$ = [-0.27, 0.40], $P_{(meaningful)}$ = 43%).

GARS-3 scores were available for 29.5% of children who received ASD diagnoses (n = 46; n females = 26) and 16.7% of non-ASD children (n = 11; n females = 3). A Bayesian ANOVA revealed differences in GARS-3 index scores between some of the four groups (Figure 4.2, right panel). Once again, there was a large difference in scores between females who did (M = 92.09, SD = 14.60) and did not (M = 69.70, SD = 6.51) receive ASD diagnoses

(d = 1.40, $HDI_{80\%}$ = [0.87, 1.93], $P_{(meaningful)}$ = 99.8%). The same pattern was found in males who did (M = 100.51, SD = 17.69) and did not (M = 71.66, SD = 14.98) receive ASD diagnoses (d = 1.51, $HDI_{80\%}$ = [1.08, 1.92], $P_{(meaningful)}$ = 100%). No meaningful differences were found between males and females who did not receive ASD diagnoses (d = 0.11, $HDI_{80\%}$ = [-0.40, 0.62], $P_{(meaningful)}$ = 51.0%), but there was weak evidence that males with ASD achieved higher scores than females with ASD (d = 0.23, $HDI_{80\%}$ = [-0.06, 0.52], $P_{(meaningful)}$ = 71.1%).

Levels of Concern Raised

Table 4.13 presents the proportion of ASD behaviours for which concern was reported for each group of children by each report source. The data were considered separately by report source and ASD domain. [42]

Table 4.13

Proportion of ASD Behaviours for Which Concern Was Reported

		Proportion of Behaviours with Concern Raised Per Group (%)			
		ASD		Non-ASD	
Source	Domain	Males	Females	Males	Females
Parent report	SC	46.4	47.1	22.5	29.5
	RRBI	42.8	40.9	17.0	22.6
Diagnostician observation	SC	38.8	33.2	16.9	14.6
	RRBI	21.3	15.9	7.3	6.3
Teacher report	SC	63.0	40.8	36.6	30.6
	RRBI	32.2	16.8	16.4	13.1

Note. SC = social communication; RRBI = restricted and repetitive behaviours and interests.

[42] This data is presented in graphic form in Appendix F (Figures F.4-F.7).

Sex/gender differences in the probability that concern was raised by each source was investigated using Bayesian hierarchical conditional logistic regressions. Table 4.14 presents the main effects (effect of assessment result and sex/gender), their interaction, and simple main effects (male minus female) for the ASD and non-ASD groups. As expected, children who received ASD diagnoses were more likely to have concern raised across all domains and sources of reporting. The difference in the probability of concern was greatest in teacher report, with ASD males having a higher probability of teachers raising concern than ASD females across both domains. Parent-reported concern was more strongly associated with ASD result for males than females, and there was weaker evidence of this for teacher report.

Across both social communication and RRBIs, parents were approximately equally likely to report concern for males and females who received ASD diagnoses. However, for non-ASD children, parents of females were more likely to raise concern for their daughters in both domains than parents of males. There was weak evidence of a higher probability that diagnosticians would raise concern for males than females in both the social communication and RRBI domains, regardless of ASD result. This was also the case for teacher reported RRBIs. However, there was strong evidence that, among children diagnosed with ASD, teachers were more likely to raise concern for males across both domains.

Table 4.14

Results of Logistic Regressions for the Proportion of Concern Levels by Sex/Gender and Assessment According to Each Source

		Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		ASD: M-F		Non-ASD: M-F	
Source: Domain		LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaningful)
Parent	SC	**1.00 [0.90, 1.10]**	**1.00**	**-0.24 [-0.34, -0.14]**	**-.96**	**0.34 [0.14, 0.54]**	**.94**	-0.07 [-0.17, 0.03]	-.34	**-0.41 [-0.58, -0.23**	**-.99**
	RRBI	**1.15 [1.02, 1.27]**	**1.00**	-0.19 [-0.31, -0.06]	-.81	**0.37 [0.11, 0.61]**	**.91**	-0.00 [-0.12, 0.11]	-.14	**-0.37 [-0.60, -0.16]**	**-.94**
Diagnost. obs.	SC	**1.15 [1.00, 1.30**	**1.00**	0.19 [0.04, 0.34]	.77	-0.07 [-0.38, 0.23]	-.45	0.15 [0.01, 0.28]	.69	0.22 [-0.05, 0.49]	.71
	RRBI	**1.39 [1.02, 1.74]**	**1.00**	0.40 [0.04, 0.75]	.86	-0.37 [-1.07, 0.32]	-.70	0.20 [-0.04, 0.46]	.71	0.58 [-0.12, 1.19]	.83
Teacher	SC	**0.82 [0.60, 1.03]**	**1.00**	**0.57 [0.36, 0.79]**	**1.00**	0.52 [0.09, 0.96]	.89	**0.83 [0.61, 1.06]**	**1.00**	0.32 [-0.07, 0.67]	.77
	RRBI	**0.75 [0.46, 1.02]**	**1.00**	**0.64 [0.36, 0.93]**	**.99**	0.43 [-0.12, 0.99]	.77	**0.85 [0.58, 1.11]**	**1.00**	0.42 [-0.05, 0.92]	.81

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. SC = social communication, RRBI = repetitive and restricted behaviours and interests, LOR = log odds ratio.

Criterion Fulfillment

The probability that each criterion was met (compared to partially met or not met) was analysed using conditional logistic regression, with sex/gender and assessment decision (i.e., assessment result) specified as predictor variables. The results of the conditional logistic regressions for criterion fulfillment are presented in Table 4.15.

Table 4.15

Logistic Regression Effect of Assessment Result, Sex/Gender and Their Interaction on Criteria Met

	Effect of Ax. Result	Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction	
Criterion	LOR [$HDI_{80\%}$] [a]	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$	LOR [$HDI_{80\%}$]	$P_{(meaningful)}$
Domain A	9.48 [6.71, 12.01]	0.14 [-2.11, 2.29]	.51	-0.14 [-4.37, 4.18]	-.51
A1	7.67 [4.79, 10.62]	0.57 [-0.38, 1.70]	.72	-0.58 [-2.90, 0.92]	-.65
A2	8.32 [4.90, 12.39]	0.01 [-0.72, 0.84]	.40	-0.01 [-1.47, 1.39]	-.43
A3	7.83 [4.36, 11.67]	0.02 [-0.72, 0.84]	.43	-0.04 [-1.56, 1.32]	-.46
Domain B	8.68 [6.05, 11.24]	-0.07 [-2.09, 1.95]	-.49	-0.03 [-3.99, 3.93]	-.49
B1	1.63 [1.21, 2.08]	0.22 [-0.12, 0.62]	.65	-0.01 [-0.65, 0.63]	-.40
B2	2.49 [2.02, 2.95]	-0.21 [-0.64, 0.16]	-.63	-0.05 [-0.77, 0.58]	-.46
B3	2.01 [1.47, 2.51]	0.62 [0.06, 1.10]	.91	-0.03 [-0.78, 0.72]	-.44
B4	3.12 [2.59, 3.65]	-0.40 [-0.88, 0.10]	-.80	-0.01 [-0.74, 0.71]	-.41

Note. Criterion A1: deficits in social-emotional reciprocity; A2: deficits in nonverbal communication behaviours; A3: deficits in developing, maintaining, and understanding relationships; B1: stereotyped/ repetitive motor movements, use of objects, or speech; B2: insistence on sameness, routines, or ritualised behaviour; B3: restricted and fixated interests; B4: hyper-/hypo-reactivity to sensory input. Positive LORs (ASD assessment result) = greater probability of being met if the result was positive for ASD; positive LORs (sex/gender) = greater probability of being met for males. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

[a] For each criterion, there was a 1.00 probability of the LOR posterior falling outside the ROPE.

As expected, there was a higher probability that a child would meet any single ASD criterion if they received an ASD diagnosis. This effect was strongest in the social communication criteria (Domain A) because, if ASD diagnosis is provided, all Domain A criteria must be met. As this is not the case for the Domain B (RRBI criteria), there was variability as to how strongly meeting a Domain B criterion predicted overall diagnosis: a diagnosis of ASD was most likely if B2 or B4 was met. Females were weakly more likely to meet Criteria B2 and B4 than were males.

The proportions of children meeting and showing *any* difficulty in each criterion (i.e., criterion met or partially met) were estimated from the logistic regression models and are presented in Table 4.16, with relative risks (proportion of ASD males or females - proportion of non-ASD males/females). Figure 4.3 visually presents the frequencies of criteria met, partially met and not met by sex/gender and assessment result.

According to the proportions outlined in Table 4.16, most non-ASD females partially (61.0%) or fully met (26.7%) Criterion A1 (social and emotional reciprocity), and only 12.3% showed no atypicality. In contrast, a smaller proportion of non-ASD males partially met A1 (33.4%), but a greater proportion fully met the criterion (51.1%). That is, more non-ASD females showed some difficulty in this area but not enough to meet the criterion compared to males, who more commonly met the criterion. With regard to nonverbal communication (A2), an approximately equal proportion of non-ASD males and females met the criterion (~30% of each), but there was a slightly larger proportion of non-ASD females who demonstrated any difficulty compared to males (77.0% and 69.2%, respectively). Only one non-ASD female was deemed to demonstrate no difficulty in relationships (A3), compared to 19.3% of non-ASD males. Non-ASD males and females had approximately the same probability of meeting this criterion (43.1% and 40.4%, respectively). Given that

Table 4.16

Descriptive Statistics and Relative Risk for Criteria Met/Partially Met by Sex/Gender and Assessment Result

Variable		Criterion						
		A1	A2	A3	B1	B2	B3	B4
Males								
Proportion meeting criterion (%)	ASD	100.0	100.0	100.0	68.8	81.8	67.8	89.9
	Non-ASD	51.1	30.0	43.1	30.6	28.2	22.8	28.7
	Relative risk	1.96	**3.33**	2.32	2.25	2.90	2.97	3.13
Proportion with any atypicality (%)	ASD	100.0	100.0	100.0	93.8	93.8	87.6	97.2
	Non-ASD	84.5	69.2	80.7	57.4	71.4	56.4	76.2
	Relative risk	1.18	1.45	1.24	**1.63**	1.31	1.55	1.28
Females								
Proportion meeting criterion (%)	ASD	100.0	100.0	100.0	63.6	85.7	53.9	93.2
	Non-ASD	26.7	29.3	40.4	25.6	32.2	13.3	37.6
	Relative risk	3.75	3.41	2.48	2.48	2.66	**4.05**	2.48
Proportion with any atypicality (%)	ASD	100.0	100.0	100.0	89.0	93.4	83.0	98.1
	Non-ASD	87.7	77.0	92.1	61.2	75.2	34.8	72.7
	Relative risk	1.14	1.30	1.09	1.45	1.24	**2.38**	1.35

Note. Proportion with any atypicality indicates the proportion of children who met or partially met the criterion. All proportions were derived from the logistic regression models. Relative risk was calculated as follows: Proportion ASD ÷ proportion non-ASD. The largest relative risk in each set of results is highlighted in boldface.

nearly all females demonstrated at least some atypicality in Criterion A3, difficulties with relationships may be a key reason that many females present for assessment, regardless of assessment result.

Across the three social communication criteria, the most common rating was *Partially met* for non-ASD females (56.7% of ratings). This is substantially higher than the probability of a *Partially met* rating for non-ASD males (36.3%). Autistic females were only slightly

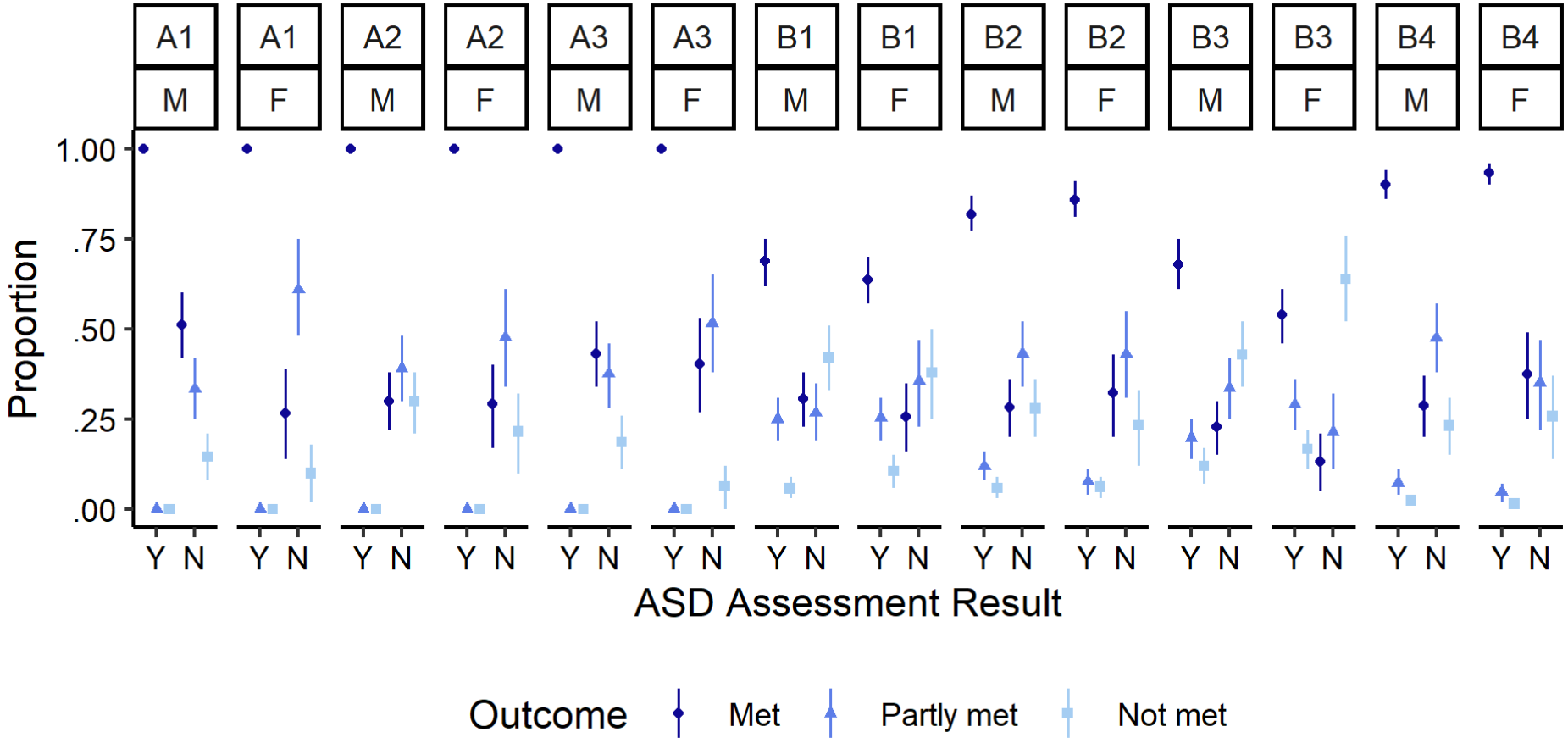
more likely to present with any atypicality than non-ASD females in each of the social communication criteria, but they were considerably more likely to fully meet each criterion. Broadly, it appears that for non-ASD females, either (a) these difficulties were not pervasive enough to warrant a criterion met (61.0%, 47.7%, and 51.7% respectively), and/or (b) they were not accompanied by sufficient atypicality in other areas.

Seventy-five percent of non-ASD females showed at least some difficulty in B2 and 72.7% in B4. While a small majority (61.2%) demonstrated some overall atypicality in B1, only 34.8% did so in B3. The order of most to least common RRBI criteria met was identical for diagnosed males and females, and non-ASD females: B4 > B2 > B1 > B3. However, non-ASD males presented a different pattern: B1 > B2 = B4 > B3. There was evidence of asymmetry in the probability of a male or female meeting two RRBI criteria, independent of their diagnostic result. Males had a higher probability of meeting Criterion B3 (restricted interests; non-ASD males had a 1.7 times higher probability than non-ASD females, and autistic males had a 1.3 times higher probability than autistic females). In contrast, there was a slightly higher probability that a female would meet Criterion B4 than a male (sensory hyper-/hypo-reactivity; non-ASD females had a 1.3 times higher probability than non-ASD males, but autistic males and females had a similar probability of meeting the criterion).

Females diagnosed with ASD were considerably more likely than non-ASD females to present with any atypicality in each of the following criteria: 1.4 times in B1, 1.2 times in B2, 3.4 times in B3, and 1.4 times in B4, and autistic females were far more likely to fully meet each criterion (3.0 times, 2.5 times, 5.6 times, and 2.3 times, respectively). Non-ASD females therefore did not show sufficient atypicality in any RRBI criterion to warrant diagnosis, but particularly so in the area of restricted interests and stereotypical behaviour.

Figure 4.3

Proportion of Criteria Met, Partially Met and Not Met by Sex/Gender and Assessment Result



Note. Error bars represent HDIs $_{(80\%)}$. M = male, F = female. Y = ASD diagnosis given, N = ASD diagnosis not given.

Social Communication

Various difficulties categorised under each social communication criterion were examined using separate logistic regressions. Consistent with the analyses presented above, sex/gender and assessment result were specified as predictor variables. Importantly, any combination of these difficulties could be considered grounds for the criterion being met – it is not necessary for an individual to display every behaviour to meet a given criterion. Thus, it is possible for some difficulties to be equally common in the ASD and non-ASD group. Identifying difficulties that are less strongly associated with a positive ASD result (i.e., ASD diagnosis being provided) will illustrate which behaviours are not generally implicated in ASD or are not associated with ASD diagnosis. Understanding how these patterns differ between males and females could further clarify whether females present with different ASD characteristics. Finally, analysis of the interaction between sex/gender and assessment result may illustrate whether an ASD diagnosis is more strongly associated with the behaviour for males or females (i.e., if a behaviour is more *diagnostic* for males or females).

For the majority of behaviours examined, the probability that atypicality was reported was higher for children who received an ASD diagnosis. However, there were some exceptions to this, particularly in areas that might be less strongly associated with ASD (e.g., difficulty losing a game).

Social-emotional reciprocity. Table 4.17 presents the results of logistic regressions for behaviours within Criterion A1 (difficulties with social-emotional reciprocity).[43] Main

[43] Refer to Appendix G (Table G.1) for the proportions of participants for whom concern was raised within each behaviour.

Table 4.17

Logistic Regression Predicting Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion A1

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Parent report											
Social approach	**1.21 [0.78, 1.63]**	**1.00**	-0.23 [-0.66, 0.20]	-.65	0.86 [-0.01, 1.71]	.87	.37 .18	0.20 [-0.29, 0.71]	.60	-0.66 [-1.35, 0.05]	-.85 *
Social norms	**1.17 [0.74, 1.59]**	**1.00**	-0.33 [-0.76, 0.11]	-.75	0.62 [-0.22, 1.49]	.78	.34 .21	-0.01 [-0.46, 0.43]	-.40	-0.64 [-1.37, 0.10]	-.82 *
Reciprocal conversation	**1.53 [1.08, 1.99]**	**1.00**	-0.11 [-0.57, 0.34]	-.51	-0.61 [-1.52, 0.30]	-.77	.29 .42	-0.42 [-0.88, 0.02]	-.82 *	0.20 [-0.63, 0.97]	.56
Sharing interests	**1.24 [0.64, 1.82]**	**1.00**	-0.66 [-1.26, -0.07]	-.89	1.23 [0.06, 2.40]	.90	.23 .11	-0.05 [-0.52, 0.43]	-.44	**-1.28 [-2.34, -0.19]**	**-.92 ***
Sharing emotions	**0.62 [0.10, 1.14]**	**.91**	-0.43 [-0.96, 0.09]	-.79	1.11 [0.09, 2.16]	.90	.17 .01	0.13 [-0.38, 0.61]	.53	-0.99 [-1.87, -0.04]	-.89 *
Content of conversation	**1.67 [1.22, 2.11]**	**1.00**	-0.00 [-0.44, 0.45]	-.39	0.34 [-0.53, 1.22]	.64	.43 .36	0.17 [-0.30, 0.65]	.57	-0.17 [-0.90, 0.58]	-.55
Literal language	-0.13 [-0.60, 0.34]	-.53	**-0.70 [-1.17, -0.23]**	**-.95**	0.18 [-0.74, 1.12]	.54	-.01 -.05	-0.60 [-1.11, -0.08]	-.90	-0.78 [-1.56, -0.00]	-.87
Diagnostic Observations											
Social approach	**1.60 [1.02, 2.16]**	**1.00**	-0.15 [-0.73, 0.41]	-.54	0.35 [-0.76, 1.52]	.61	.29 .25	0.03 [-0.41, 0.49]	.42	-0.33 [-1.38, 0.72]	-.61
Social norms	**1.13 [0.59, 1.65]**	**1.00**	0.42 [-0.12, 0.95]	.79	-0.23 [-1.29, 0.84]	-.56	.21 .21	0.30 [-0.16, 0.75]	.72	0.53 [-0.46, 1.45]	.72

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Reciprocal conversation	**1.82 [1.36, 2.28]**	**1.00**	0.53 [0.06, 1.00]	.89	0.42 [-0.50, 1.34]	.67	.46 .37	**0.74 [0.27, 1.22]**	**.96 ***	0.32 [-0.49, 1.10]	.64
Sharing interests	**2.43 [0.83, 3.97]**	**1.00**	**1.91 [0.35, 3.45]**	**.97**	-2.97 [-5.92, 0.19]	-.93	.09 .11	0.43 [-0.19, 1.04]	.76	0.32 [-0.49, 1.10]	.64
Sharing emotions	**1.05 [0.15, 1.91]**	**.94**	0.18 [-0.71, 1.05]	.55	-0.65 [-2.40, 1.11]	-.66	.05 .08	-0.14 [-0.82, 0.49]	-.53	0.49 [-1.13, 2.11]	.63
Content of conversation	**2.31 [1.57, 3.02]**	**1.00**	**1.02 [0.28, 1.74]**	**.97**	**-2.24 [-3.69, -0.77]**	**-.99**	**.28 .53**	-0.10 [-0.55, 0.34]	-.51	**2.16 [0.75, 3.54]**	**.99 ***
Literal language	**-0.93 [-1.57, -0.29]**	**-.95**	0.69 [0.01, 1.36]	.88	0.31 [-0.97, 1.60]	.58	-.09 -.07	0.86 [0.03, 1.74]	.88	0.53 [-0.48, 1.51]	.72
Teacher report											
Academic achievement	-0.04 [-0.61, 0.56]	-.44	-0.62 [-1.21, -0.03]	-.87	0.11 [-1.04, 1.27]	.51	.00 -.02	-0.56 [-1.20, 0.05]	-.83	-0.68 [-1.67, 0.32]	-.77
Social approach	0.33 [-0.20, 0.83]	.71	0.50 [-0.03, 1.01]	.84	0.66 [-0.36, 1.69]	.76	.16 .00	**0.82 [0.28, 1.37]**	**.96 ***	0.16 [-0.70, 1.03]	.53
Reciprocal conversation	**1.05 [0.49, 1.62]**	**.99**	**0.87 [0.29, 1.43]**	**.97**	0.63 [-0.50, 1.76]	.72	.33 .15	**1.18 [0.61, 1.74]**	**.99 ***	0.55 [-0.45, 1.51]	.73

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

effects of assessment result, sex/gender, and their interaction on the behaviour, [44] and simple main effects (ASD M - F; non-ASD M - F) are provided. For most of the areas examined, there was no conclusive evidence that sex/gender predicted the probability that atypicality was reported. However, there was a higher probability that parents of females would report difficulties with *Literal [interpretation of] language*.

In most cases where there was evidence of an interaction between sex/gender and assessment result (five of the behaviours assessed; $P_{(meaningful)} > .85$), that interaction was due to a large sex/gender difference for one level of diagnosis (i.e., ASD or non-ASD) accompanied by a negligible difference for the other level (particularly in *Sharing Interests* and *Sharing Emotions*). The differences tended not to show diagnostic levels with evidence of an effect in the same direction, but with different magnitudes.

Of behaviours observed by diagnosticians, there was evidence that males had a higher probability than females of demonstrating difficulty with *Social norms* (e.g., awareness of boundaries/personal space, saying socially inappropriate things), *Reciprocal conversation*, and *Literal language*. For *Conversation content,* the meaningful effect of sex/gender was mostly driven by a large difference in non-ASD children (males more likely than females to have difficulty observed). In contrast, for *Reciprocal conversation,* the overall difference was primarily carried by ASD children. Teachers were meaningfully more likely to raise concern about *Reciprocal conversation* for males and more weakly for *Social approach* (i.e., a negligible sex/gender difference, nor a difference of zero, could be excluded with 80% confidence). However, these sex/gender differences were predominately carried by the ASD

[44] The interaction terms are described using relative risks in the text. However, these were not always interpretable as denominators of the proportions sometimes differed considerably between males and females, making the relative risks difficult to compare. In Table 4.17 (and all tables thereafter), the differences in proportions are reported to assist in interpreting the interaction terms.

children. There was moderate evidence of a higher probability of teachers reporting concern about *Academic achievement* for females compared to males.

Evidence indicated interactions between assessment result and sex/gender for several behaviours within Criterion A1. For three behaviours, there was moderate evidence that assessment result (i.e., receiving an ASD diagnosis compared to not receiving an ASD diagnosis) was more strongly predictive of the presence of the behaviour for males than females: parent-reported difficulties in *Social approach* (relative risk for males; $RR_M = 2.0$, relative risk for females; $RR_F = 1.3$), *Sharing interests* ($RR_M = 4.8$, $RR_F = 1.3$), and *Sharing emotions* ($RR_M = 2.7$, $RR_F = 1.0$). Indeed, there was only a small difference between ASD and non-ASD females in the probability that these behaviours would be reported. Finally, the association between assessment result and the presence of two diagnostic observations was stronger for males: *Sharing interests* and *Content of conversation* (refer to Table 4.17).

Nonverbal Communication. Logistic regressions were applied to assess how strongly the presence of specific nonverbal communication difficulties were predicted by assessment result and sex/gender (refer to Table 4.18). The probability that a child would present with any difficulty in Criterion A2 was meaningfully higher if they received an ASD diagnosis, with the exception of teacher-reported difficulties using *Body language* (where a negligible difference could not be excluded with 80% confidence). The effect of sex/gender was not universal and there was no conclusive evidence of an association between sex/gender and approximately half of the behaviours examined. However, females had a higher probability of having parent-reported difficulties making consistent *Eye contact, Interpreting nonverbal behaviour,* and using appropriate *Facial expressions*. For the latter, evidence of a sex/gender difference was strong, and a negligible difference could be excluded with 80% confidence. Atypical use of *Facial expression* was also more likely to be observed by

diagnosticians among females than males. In contrast, the probability of teacher-reported concern in both A2 difficulties was higher for males.

There was evidence of an interaction between assessment result and sex/gender ($P_{(meaningful)} \geq 80\%$) in only three behaviours: *Nonverbal understanding* (teacher report; $RR_M = 2.4$, $RR_F = 1.6$), *Use of nonverbal communication* (parent report; $RR_M = 3.0$, $RR_F = 1.3$), and *Integration of verbal and nonverbal behaviour* (parent report; refer to Table 4.18). In these areas, there was a larger increase in probability of reporting atypicality for males than females if they received ASD diagnosis.

Table 4.18

Logistic Regression Predicting Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion A2

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Parent report											
Integration of verbal/NV behaviour	**3.80 [1.93, 5.49]**	**1.00**	**-2.08 [-3.73, -0.42]**	**-.96**	2.27 [-1.03, 5.50]	.83	.14 .26	**-0.93 [-1.49, -0.40]**	**-.98**	**-3.32 [-6.68, -0.14]**	**-.93**
Eye contact	**1.06 [0.64, 1.49]**	**1.00**	-0.49 [-0.92, -0.06]	-.88	0.30 [-0.55, 1.16]	.62	.28 .22	-0.34 [-0.77, 0.11]	-.76	-0.64 [-1.40, 0.08]	-.83
Use of nonverbal comm.	**0.86 [0.32, 1.39]**	**.97**	-0.24 [-0.80, 0.29]	-.63	0.93 [-0.14, 2.01]	.84	.19 .07	0.22 [-0.26, 0.71]	.63	-0.71 [-1.68, 0.23]	-.79
Facial expression	**1.54 [1.04, 2.04]**	**1.00**	**-1.17 [-1.67, -0.67]**	**-1.00**	0.19 [-0.79, 1.19]	.54	.27 .34	**-1.07 [-1.52, -0.62]**	**-1.00**	**-1.26 [-2.15, -0.37]**	**-.95**
Nonverbal understand.	**1.30 [0.87, 1.73]**	**1.00**	-0.53 [-0.95, -0.09]	-.90	0.39 [-0.48, 1.24]	.67	.35 .26	-0.34 [-0.79, 0.13]	-.74	-0.72 [-1.44, 0.02]	-.86
Response to NV bhvr.	**1.41 [0.89, 1.91]**	**1.00**	-0.22 [-0.72, 0.29]	-.62	0.16 [-0.86, 1.15]	.53	.28 .28	-0.14 [-0.58, 0.30]	-.55	-0.30 [-1.22, 0.59]	-.61
Emotional regulation	**0.80 [0.37, 1.21]**	**.99**	0.16 [-0.27, 0.58]	.57	-0.27 [-1.10, 0.57]	-.60	.16 .23	0.02 [-0.42, 0.46]	.41	0.30 [-0.43, 1.01]	-.63
Diagnostic observations											
Eye contact	**1.35 [0.84, 1.83]**	**1.00**	0.01 [-0.49, 0.51]	.41	0.30 [-0.67, 1.32]	.61	.30 .24	0.16 [-0.28, 0.59]	.57	-0.14 [-1.04, 0.75]	-.52

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Use of nonverbal comm.	**1.13 [0.52, 1.71]**	**.99**	-0.13 [-0.73, 0.47]	-.53	0.86 [-0.31, 2.06]	.80	.21 .10	0.30 [-0.20, 0.78]	.70	-0.56 [-1.65, 0.51]	-.71
Facial expression	**1.14 [0.65, 1.61]**	**1.00**	-0.57 [-1.06, -0.08]	-.89	0.07 [-0.86, 1.06]	.49	.21 .25	-0.53 [-0.97, -0.08]	-.89	-0.61 [-1.47, 0.26]	-.77
Nonverbal understand.	**1.12 [0.64, 1.57]**	**1.00**	0.12 [-0.35, 0.59]	.52	0.57 [-0.35, 1.51]	.74	.31 .18	0.40 [-0.04, 0.84]	.81	-0.17 [-0.99, 0.65]	-.54
Teacher report											
Use of nonverbal comm.	0.65 [0.01, 1.25]	.89	**0.99 [0.34, 1.60]**	**.96**	-0.53 [-1.77, 0.72]	-.68	.09 .15	**0.72 [0.13, 1.27]**	**.92**	**1.25 [0.13, 2.35]**	**.92**
Nonverbal understand.	**1.30 [0.70, 1.88]**	**1.00**	0.61 [0.02, 1.20]	.87	1.07 [-0.09, 2.28]	.85	.43 .17	**1.14 [0.57, 1.73]**	**.99 ***	0.07 [-0.94, 1.10]	.49

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

Developing, Maintaining, and Understanding Relationships. Logistic regressions were used to assess how strongly the presence of specific difficulties with relationships were predicted by ASD assessment result and sex/gender (Table 4.19). The probability that a child presented with all but two Criterion A3 difficulties was meaningfully higher if they received an ASD diagnosis. This illustrates that most A3 difficulties were associated with a positive ASD result, but that some were present in the referred sample but not associated with ASD diagnosis. For example, in *Social motivation*, or an individual's interest in having friends and being social, neither assessment result nor sex/gender meaningfully predicted its excessiveness or absence. Although there was equivocal evidence of an overall sex/gender difference in *Possessiveness and [difficulty] losing [games]* and *Friendship formation* according to parent report, ASD males were meaningfully more likely to have concern reported than ASD females. For *Friendship formation*, there was probable evidence that non-ASD females were more likely to have concern raised than non-ASD males. This was also seen in diagnostician observed difficulties in *Imaginative/spontaneous play.*

Moderate evidence was found to suggest a higher probability of parent-reported difficulty in *Imaginative/spontaneous play* and a higher probability of teacher-reported difficulty with *Friendship formation* for males compared to females. In contrast, there was strong evidence of a higher probability of parents reporting difficulties with overly *Submissive or dominating [behaviour] in play* for females compared to males.

Interactions between assessment result and sex/gender were found in all diagnostic observations examined within Criterion A3. Atypicality in *Imaginative/spontaneous play* was less strongly predicted by positive ASD result for females than males ($RR_M = 5.3$, $RR_F =$

Table 4.19

Logistic Regression Predicting Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion A3

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Parent report											
Adjusting behaviour for situation	**1.38 [0.73, 2.00]**	**1.00**	0.07 [-0.57, 0.69]	.48	-0.21 [-1.46, 1.05]	-.55	.19 .21	-0.04 [-0.52, 0.44]	-.43	0.18 [-0.99, 1.35]	.53
Imaginative play	**0.88 [0.35, 1.38]**	**.98**	0.54 [0.02, 1.06]	.87	-0.43 [-1.44, 0.64]	-.66	.14 .18	0.33 [-0.11, 0.80]	.74	0.75 [-0.23, 1.66]	.83
Submissive/ dominating in play	0.28 [-0.13, 0.71]	.71	**-0.69 [-1.11, -0.25]**	**-.96**	0.14 [-0.70, 0.99]	.53	.08 .05	**-0.61 [-1.07, -0.18]**	**-.93**	-0.76 [-1.46, -0.01]	-.88
Possessive/ losing	**0.80 [0.38, 1.23]**	**.98**	0.19 [-0.24, 0.62]	.61	0.79 [-0.07, 1.65]	.85	.28 .09	**0.58 [0.15, 1.03]**	**.92 ***	-0.22 [-0.94, 0.53]	-.57
Friendship formation	**1.44 [1.01, 1.88]**	**1.00**	-0.07 [-0.51, 0.37]	-.47	**1.39 [0.53, 2.27]**	**.97**	**.48 .18**	**0.62 [0.16, 1.08]**	**.93 ***	-0.76 [-1.49, -0.00]	-.87
Friendship maintenance	**1.41 [0.98, 1.83]**	**1.00**	-0.17 [-0.59, 0.27]	-.58	0.51 [-0.34, 1.40]	.73	.39 .27	0.09 [-0.39, 0.57]	.49	-0.43 [-1.14, 0.27]	-.72
Social motivation	0.15 [-0.37, 0.67]	.55	0.12 [-0.40, 0.64]	.52	0.19 [-0.83, 1.25]	.55	.04 .01	0.22 [-0.31, 0.75]	.61	0.02 [-0.90, 0.89]	.46
Consistent companions	**0.68 [0.21, 1.13]**	**.95**	0.18 [-0.29, 0.64]	.59	-0.04 [-0.94, 0.90]	-.47	.18 .14	0.16 [-0.29, 0.61]	.57	0.19 [-0.60, 1.01]	.56

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Diagnostic observations											
Friendship understand.	**1.01 [0.58, 1.45]**	**1.00**	0.05 [-0.41, 0.49]	.44	-0.74 [-1.63, 0.16,]	-.82	.15 .31	-0.32 [-0.75, 0.12]	-.74	0.41 [-0.37, 1.21]	.70
Inclusiveness in play	**1.79 [0.95, 2.60]**	**1.00**	0.38 [-0.47, 1.17]	.67	**-1.85 [-3.46, -0.18]**	**-.93**	**.08 .22**	-0.55 [-1.11, -0.02]	-.85 *	1.29 [-0.23, 2.84]	.86 *
Imaginative/ spont. play	**1.38 [0.56, 2.17]**	**.99**	0.06 [-0.74, 0.86]	.47	**1.88 [0.29, 3.47]**	**.92**	**.39 .07**	**0.99 [0.39, 1.60]**	**.97 ***	-0.89 [-2.32, 0.63]	-.76
Teacher report											
Friendship formation	**0.99 [0.44, 1.51]**	**.99**	0.55 [0.01, 1.09]	.86	0.43 [-0.63, 1.52]	.65	.29 .17	**0.76 [0.21, 1.33]**	**.94 ***	0.34 [-0.59, 1.24]	.63
Friendship maintenance	**0.96 [0.44, 1.48]**	**.98**	0.25 [-0.28, 0.77]	.64	0.60 [-0.44, 1.66]	.73	.30 .16	0.55 [-0.02, 1.10]	.85	-0.05 [-0.95, 0.81]	-.47

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

1.4). A similar pattern was observed for parent-reported difficulties with *Possessiveness and losing games* (RR_M = 1.9, RR_F = 1.3) and *Friendship formation* (RR_M = 3.0, RR_F = 1.5). In contrast, diagnostician observation of difficulties with *Understanding friendships* (RR_M = 1.4, RR_F = 2.3) and *Inclusiveness in play* (refer to Table 4.19) were more strongly predictive of a positive ASD result for females than males.

Repetitive and Restricted Behaviours and Interests

Stereotypical and Repetitive Behaviour. Logistic regressions were used to examine how strongly sex/gender and assessment result (predictor variables) were associated with the presence of specific stereotypical behaviours (outcome variables; Table 4.20). Many behaviours were predicted by ASD result, and additionally, sex/gender was a meaningful predictor of whether certain behaviours were reported by parents, independent of assessment result.

Within the motor stereotypes domain, there was moderate evidence of a higher probability of *Repetitive body use* behaviours (such as nail biting or hair twirling) among females. However, for *Toe walking* and *Spinning* behaviours, there was only a meaningful difference in probability that females, and not males, would demonstrate the behaviour if they received an ASD diagnosis (*Toe walking*: RR_M = 0.83, RR_F = 4.8; *Spinning:* RR_M = 0.9, RR_F = 4.0).[45] The interaction between assessment result and sex/gender in *Rocking/jumping* behaviours indicated a meaningful difference only for males (RR_M = 2.7, RR_F = 1.1). [46, 47]

[45] A relative risk *less than* 1.0 indicates that the probability of the behaviour being reported was higher for non-ASD children than for ASD children.

[46] Data were also collected for parent reported *Mouth mannerisms* (e.g., grimacing or mouth posturing). However, this mannerism was not reported frequently, and therefore the results were not interpretable. The results of this logistic regression are therefore not reported here but are included with other excluded variables in Appendix G (Table G.2).

[47] The frequency of each stereotypic behaviour by sex/gender and assessment result is presented in graphic form in Appendix H (Figures H.1 to H.3).

Table 4.20

Logistic Regression Predicting Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion B1, Parent Report

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Motor stereotypies	1.10 [0.58, 1.61]	**1.00**	0.34 [-0.19, 0.84]	.73	-0.82 [-1.82, 0.24]	-.82	.15 .28	-0.07 [-0.51, 0.38]	-.47	0.73 [-0.20, 1.65]	.82
Toe walking	0.79 [-0.01, 1.54]	.90	0.76 [-0.03, 1.53]	.89	**-1.96 [-3.48, -0.38]**	**-.96**	**-.03 .15**	-0.22 [-0.77, 0.36]	-.61	**1.74 [0.27, 3.15]**	**.96 ***
Flapping	0.76 [0.09, 1.38]	.92	-0.07 [-0.75, 0.56]	-.48	-0.51 [-1.66, 0.81]	-.66	.06 .12	-0.33 [-0.89, 0.23]	-.70	0.17 [-1.05, 1.30]	.53
Spinning	0.73 [-0.13, 1.57]	.85	0.07 [-0.80, 0.92]	.48	-1.59 [-0.80, 0.18]	-.88	-.01 .09	-0.72 [-1.39, 0.00]	-.88 *	0.85 [-0.75, 2.39]	.74
Gross motor mannerism	-0.14 [-0.71, 0.39]	-.54	-0.34 [-0.91, 0.22]	-.71	-0.15 [-1.27, 0.94]	-.52	-.02 -.01	-0.41 [-1.02, 0.22]	-.74	-0.27 [-1.21, 0.64]	-.59
Rocking/ jumping	**0.69 [0.13, 1.23]**	**.92**	-0.12 [-0.69, 0.43]	-.52	1.03 [-0.03, 2.15]	.87	.09 .02	0.39 [-0.13, 0.91]	.77	-0.65 [-1.61, 0.35]	-.76
Rigidity	0.65 [-0.06, 1.31]	.86	-0.35 [-1.04, 0.34]	-.68	-0.51 [-1.88, 0.85]	-.66	.03 .09	-0.59 [-1.22, 0.00]	-.85 *	-0.09 [-1.32, 1.12]	-.50
Hand mannerisms	0.04 [-0.46, 0.59]	.44	0.11 [-0.42, 0.64]	.51	0.50 [-0.58, 1.53]	.69	.04 -.03	0.35 [-0.20, 0.92]	.72	-0.14 [-1.03, 0.75]	-.52
Self-injurious	**0.72 [0.15, 1.28]**	**.93**	0.04 [-0.53, 0.62]	.45	0.27 [-0.84, 1.43]	.57	.12 .08	0.18 [-0.35, 0.68]	.58	-0.09 [-1.11, 0.91]	-.49
Repetitive body use	**-0.84 [-1.46, -0.21]**	**-.93**	-0.65 [-1.31, -0.03]	-.87	-0.37 [-1.64, 0.85]	-.61	-.07 -.08	-0.84 [-1.66, 0.03]	-.88	-0.48 [-1.41, 0.46]	-.70

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Speech/ language	**3.38 [1.95, 4.72]**	**1.00**	**2.28 [0.87, 3.63]**	**1.00**	**-3.63 [-6.29, -0.80]**	**-.99**	**.34 .44**	0.47 [0.02, 0.91]	.86	**4.05 [1.26, 6.74]**	**1.00**
Echolalia	**0.97 [0.11, 1.75]**	**.94**	**0.95 [0.12, 1.75]**	**.93**	-0.81 [-2.42, 0.81]	-.72	.07 .08	0.54 [-0.06, 1.13]	.83	1.33 [-0.19, 2.84]	.88
Third person referencing	0.15 [-0.87, 1.15]	.53	-0.67 [-1.66, 0.34]	-.78	**1.77 [0.26, 3.67]**	**.87**	**.02 -.04**	0.19 [-0.86, 1.20]	.54	-1.55 [-3.26, 0.11]	-.88 *
Neologisms	0.43 [-0.10, 0.97]	.80	**-0.85 [-1.39, -0.31]**	**-.97**	**1.37 [0.28, 2.42]**	**.94**	**.13 -.05**	-0.17 [-0.68, 0.35]	-.57	**-1.54 [-2.46, -0.59]**	**-.98 ***
Pronoun reversal	0.40 [-0.45, 1.17]	.67	0.58 [-0.27, 1.35]	.79	-1.76 [-3.30, -0.07]	-.93	-.05 .08	-0.29 [-0.98, 0.39]	-.64	1.42 [-0.05, 2.86]	.91 *
Repetitive speech	**0.99 [0.47, 1.49]**	**.99**	**0.99 [0.47, 1.49]**	**.99**	-0.97 [-1.99, 0.05]	-.87	.12 .27	0.50 [0.05, 0.93]	.88	**1.47 [0.55, 2.38]**	**.98**
Accents	**1.14 [0.43, 1.84]**	**.98**	**-1.25 [-1.94, -0.53]**	**-.98**	0.29 [-1.11, 1.69]	.57	.09 .15	**-1.09 [-1.67, -0.54]**	**-.99**	**-1.39 [-2.67, -0.12]**	**-.91**
Unusual noises	**0.98 [0.49, 1.47]**	**.99**	0.38 [-0.11, 0.89]	.77	0.73 [-0.23, 1.73]	.79	.28 .11	**0.74 [0.30, 1.21]**	**.97 ***	0.02 [-0.89, 0.87]	.45
Talking to self	-0.17 [-1.29, 0.87]	-.53	**-1.35 [-2.42, -0.23]**	**-.94**	0.68 [-1.40, 2.82]	.64	.00 -.03	-0.98 [-2.23, 0.20]	-.84	-1.69 [-3.42, 0.04]	-.90
Odd prosody	**1.24 [0.51, 1.97]**	**.99**	**1.12 [0.36, 1.82]**	**.98**	**-2.75 [-4.15, -1.26]**	**-1.00**	**-.03 .33**	-0.25 [-0.73, 0.20]	-.66	**2.50 [1.10, 3.87]**	**1.00 ***

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Object use	1.36 [0.72, 1.99]	**1.00**	0.07 [-0.57, 0.69]	.48	-0.71 [-1.95, 0.56]	-.74	.18 .24	-0.28 [-0.76, 0.20]	-.69	0.42 [-0.72, 1.59]	.64
Lining up	-0.14 [-0.56, 0.28]	-.55	0.31 [-1.19, 0.73]	.73	0.33 [-0.50, 1.19]	.64	.01 -.07	0.47 [0.02, 0.92]	.85 *	0.14 [-0.56, 0.88]	.53
Grouping	0.07 [-0.43, 0.56]	.47	-0.45 [-0.94, 0.06]	-.82	0.01 [-0.99, 1.00]	.45	.00 .01	-0.46 [-0.98, 0.08]	-.81	-0.45 [-1.32, 0.38]	-.70
Spinning/ flicking/ pushing	**1.92 [0.96, 2.80]**	**1.00**	0.30 [-0.59, 1.20]	.62	0.55 [-1.27, 2.31]	.63	.21 .12	0.57 [0.04, 1.13]	.87 *	0.02 [-1.76, 1.67]	.48
Repetitive play	0.55 [-0.15, 1.22]	.83	-0.61 [-1.28, 0.08]	-.83	0.43 [-0.91, 1.80]	.62	.04 .03	-0.39 [-1.02, 0.23]	-.73	-0.83 [-2.03, 0.36]	-.79
Deconstruct-ion	**3.22 [1.45, 4.86]**	**1.00**	-0.78 [-2.38, 0.81]	-.72	**4.11 [0.88, 7.20]**	**.97**	**.24 .05**	-0.39 [-1.02, 0.23]	-.73	-0.83 [-2.03, 0.36]	-.79

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

There was stronger evidence of an effect of sex/gender within the speech and language domain, with a higher probability of *Echolalia* and *Repetitive speech* among males and of *Neologisms*, *Accents,* and *Talking to oneself* among females.

The interactions between assessment result and sex/gender revealed a stronger association between assessment result and the presence of the behaviour for females in *Repetitive speech* ($RR_M = 1.3$, $RR_F = 3.0$) and *Odd prosody of speech* (refer to Table 4.20). Indeed, the interactions in these cases were mostly carried by the effect among females. The differences in the probability of atypicality by assessment result were in opposite directions between males and females for *Third person referencing* ($RR_M = 0.4$, $RR_F = 2.0$) and *Pronoun reversals* ($RR_M = 0.6$, $RR_F = 3.0$). For females regarding *Third person referencing* and for males regarding *Pronoun reversals*, receiving a positive ASD result was weakly associated with a lower probability of the behaviour being reported. However, this was most apparent in the use of *Neologisms* ($RR_M = 2.6$, $RR_F = 0.8$). Broadly, parents of males were meaningfully more likely to report difficulties with *Speech/language stereotypes*, but the difference in probability by assessment result was meaningfully larger for females than males (refer to Table 4.20).

While there was no evidence that sex/gender meaningfully predicted atypical *Object use*, evidence suggested that female sex/gender was weakly associated with *Object grouping* and *Repetitive play*. There was a meaningful interaction between assessment result and sex/gender in *Deconstruction* behaviours, with a larger difference in probability of reporting this for males than females. Finally, there was weak evidence that the presence of *Motor stereotypies* and *Object use stereotypies* were more *diagnostic* (i.e., more strongly associated with positive assessment result) for females than males ($RR_M = 1.5$, $RR_F = 3.0$; and $RR_M = 2.2$, $RR_F = 4.0$, respectively).

Diagnostician observed atypicality in *Stereotypical speech/language* was strongly associated with positive ASD assessment result (refer to Table 4.21).[48] However, atypicality in *Speech/language stereotypes* was the only teacher-report domain to be meaningfully predicted by assessment result, and there was strong evidence of a higher probability of concern among males than females. Finally, there was a meaningful interaction between assessment result and sex/gender in teacher-reported *Stereotypical object use* ($RR_M = 3.0$, $RR_F = 0.1$), indicating that this was more strongly associated with assessment result for males compared to females.

[48] Data were also collected for *Stereotypical object use* (diagnostician observation). As this was not reported frequently the results were not interpretable and are therefore not reported here but presented in Appendix G (Table G.2).

Table 4.21

Logistic Regression Predicting Stereotypical Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion B1, Diagnostic Observation and Teacher Report

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Diagnostic observations											
Stereotypical movement	0.73 [-0.17, 1.56]	.84	0.53 [-0.33, 1.40]	.75	-0.61 [-2.34, 1.09]	-.66	.14 .28	0.23 [-0.47, 0.92]	.59	0.80 [-0.80, 2.31]	.74
Stereotypical speech/lang.	**1.06 [0.38, 1.72]**	**.98**	-0.06 [-0.73, 0.61]	-.47	0.00 [-1.33, 1.34]	-.46	.34 .44	-0.05 [-0.59, 0.48]	-.46	-0.08 [-1.32, 1.13]	-.49
Teacher report											
Stereotypical movement	0.43 [-0.31, 1.15]	.72	0.61 [-0.14, 1.34]	.82	1.05 [-0.42, 2.51]	.79	.14 .00	**1.12 [0.33, 1.88]**	**.96**	0.08 [-1.16, 1.33]	.49
Stereotypical speech/lang.	**1.02 [0.17, 1.81]**	**.95**	**1.20 [0.37, 2.03]**	**.97**	-0.35 [-1.97, 1.31]	-.58	.18 .12	**1.01 [0.35, 1.68]**	**.96**	**1.36 [-0.19, 2.80]**	**.89**
Stereotypical object use	-0.92 [-2.65, 0.79]	-.74	**2.49 [0.64, 4.31]**	**.98**	**4.34 [1.09, 7.72]**	**.97**	.10 -.04	**4.54 [1.66, 7.35]**	**1.00 ***	0.29 [-1.62, 2.13]	.55

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

***** Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

Insistence on Sameness, Routines and Rituals. As can be seen in Table 4.22, for all but two behavioural domains (i.e., diagnostician observed *Routine adherence* and teacher reported *Distress at change*), children who received an ASD diagnosis were meaningfully more likely to present with the behaviour, as one might expect. Evidence of a meaningful effect of sex/gender was only found for parent-reported *Routine adherence* among non-ASD children. There was weak evidence that diagnosticians were more likely to observe difficulty with *Routine adherence* for males and *Cognitive rigidity* for females, and that teachers were more likely to report *Distress at change* for males than females.[49]

For two domains within this criterion, there was evidence of a meaningful interaction between assessment result and sex/gender: parent-reported *Distress at change* ($RR_M = 2.8$, $RR_F = 1.3$) and *Cognitive rigidity* ($RR_M = 3.8$, $RR_F = 1.9$). Weaker evidence of this was found for parent-reported *Routine adherence* ($RR_M = 2.6$, $RR_F = 1.3$), but a small effect in the opposite direction could not be excluded with 80% confidence.

[49] Data were also collected for *Task switching/transitioning* (diagnostician observation) and *Routine adherence* (teacher report). Results were not interpretable as the behaviours were not reported frequently (refer to Appendix G; Table G.2).

Table 4.22

Logistic Regression Predicting Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion B2

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Parent report											
Distress at change	**1.33 [0.89, 1.75]**	**1.00**	-0.37 [-0.81, 0.07]	-.79	**1.34 [0.48, 2.21]**	**.97**	**.46 .16**	0.30 [-0.17, 0.76]	.70	**-1.04 [-1.78, -0.32]**	**-.95 ***
Routine adherence	**1.04 [0.60, 1.47]**	**1.00**	**-0.61 [-1.05, -0.16]**	**-.93**	0.83 [-0.04, 1.71]	.86	.31 .15	-0.19 [-0.62, 0.25]	-.61	**-1.03 [-1.79, -0.26]**	**-.94 ***
Task switching/ transitioning	**1.41 [0.90, 1.90]**	**1.00**	-0.27 [-0.78, 0.23]	-.66	0.09 [-0.90, 1.11]	.50	.27 .28	-0.22 [-0.66, 0.21]	-.63	-0.32 [-1.26, 0.56]	-.62
Cognitive rigidity	**1.82 [1.36, 2.27]**	**1.00**	-0.21 [-0.66, 0.26]	-.62	**1.12 [0.23, 2.04]**	**.93**	**.53 .30**	0.35 [-0.13, 0.80]	.75	-0.77 [-1.55, 0.00]	-.86 *
Diagnostic observations											
Routine adherence	0.59 [-0.52, 1.57]	.73	0.57 [-0.47, 1.60]	.73	0.21 [-1.85, 2.25]	.53	.01 .01	0.65 [-0.32, 1.60]	.78	0.48 [-1.31, 2.31]	.61
Cognitive rigidity	**1.31 [0.61, 2.00]**	**.99**	-0.48 [-1.18, 0.22]	-.76	-0.11 [-1.49, 1.27]	-.51	.11 .17	-0.54 [-1.07, -0.01]	-.86	-0.43 [-1.74, 0.82]	-.63
Teacher report											
Distress at change	0.63 [0.06, 1.20]	.89	0.44 [-0.14, 1.01]	.79	0.26 [-0.85, 1.42]	.57	.17 .10	0.57 [0.01, 1.13]	.86	0.32 [-0.67, 1.31]	.61

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

Restricted Interests. Table 4.23 presents logistic regressions for specific restricted interests according to parent report, as predicted by sex/gender and assessment result. The presence of many of these interests was associated with ASD diagnosis. However, restricted interest in *Specific characters/programs* appeared unrelated to assessment result, and surprisingly, restricted interests in *Vehicles* was inversely related to assessment result (i.e., associated with a negative ASD result). Regardless of assessment result, parents of females were more likely to report restricted interests in *Craft* and other *People.* Conversely, there was strong evidence of a higher probability that parents of males would report restricted interests in *Screens* and *Vehicles,* with the sex/gender difference in the latter primarily carried by ASD children.[50]

Three general patterns of interactions emerged from the data. The first, indicating a different strength of the effect of the association between assessment result and the restricted interest according to the child's sex/gender, was found in interests in *Animals* ($RR_M = 1.8$, $RR_F = 7.0$) and *People* (refer to Table 4.23). Second, there was a difference in the probability of restricted interests in *Toys* and *Specific programs/characters* by assessment result for females ($RR_F = 2.1$; 2.7) but not males ($RR_M = 1.1$; 1.0, respectively). In contrast, the interaction effect in restricted interest in *Random objects* was carried by males only ($RR_M = 2.7$, $RR_F = 1.0$). Finally, the association between assessment result and restricted interest in *Vehicles* differed in direction between males and females (i.e., there was a higher probability of being reported for ASD males compared to non-ASD males, $RR_M = 1.8$; but lower probability of being reported for ASD females compared to non-ASD females, $RR_F = 0.1$).[51]

---

[50] Data were also collected for restricted interests in *Different places/times* (e.g., historical periods) and *Self-presentation* (e.g., make-up, fashion). Due to low frequency of reporting, these results were uninterpretable (presented in Appendix G; Table G.2).

[51] The frequency of each restricted interest by sex/gender and assessment result is presented in graphic form in Appendix H (Figure H.4).

Table 4.23

Logistic Regression Predicting Restricted Interest by Assessment Result, Sex/Gender, and Their Interaction: Criterion B3, Parent Report

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Restricted Interest	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Specific program/ character	0.18 [-0.27, 0.62]	.60	-0.03 [-0.48, 0.43]	-.42	**-1.25 [-2.17, -0.36]**	-.95	-.01 .18	**-0.65 [-1.10, -0.18]**	**-.94**	0.59 [-0.20, 1.36]	.80
Random objects	**0.72 [0.27, 1.17]**	**.97**	-0.44 [-0.91, -0.00]	-.84	**1.43 [0.56, 2.33]**	**.97**	**.29 .00**	0.27 [-0.18, 0.70]	.69	**-1.16 [-1.92, -0.37]**	**-.96 ***
Vehicles	-0.90 [-1.67, -0.02]	-.90	**2.17 [1.31, 2.99]**	**1.00**	**3.38 [1.73, 5.02]**	**1.00**	**.15 -.12**	**3.77 [2.41, 5.00]**	**1.00 ***	0.46 [-0.51, 1.46]	.68
Toys	0.54 [0.06, 1.00]	.89	0.35 [-0.13, 0.82]	.75	-0.90 [-1.85, 0.05]	-.87	.02 .20	-0.10 [-0.55, 0.35]	-.50	0.79 [-0.06, 1.61]	.86 *
Screens	**1.49 [0.73, 2.27]**	**1.00**	**1.96 [1.15, 2.73]**	**1.00**	-0.61 [-2.07, 0.97]	-.67	.27 .14	**1.65 [1.11, 2.16]**	**1.00**	**2.25 [0.82, 3.70]**	**.99**
Animals	**1.48 [0.67, 2.25]**	**1.00**	0.44 [-0.38, 1.18]	.72	-1.50 [-2.99, 0.11]	-.90	.10 .24	-0.31 [-0.82, 0.18]	-.70	1.19 [-0.31, 2.64]	.85 *
Systems	**1.78 [0.67, 2.83]**	**.99**	-0.25 [-1.29, 0.82]	-.57	0.86 [-1.24, 2.90]	.69	.11 .08	0.18 [-0.48, 0.86]	.56	-0.75 [-2.70, 1.27]	-.67
Craft	0.72 [0.07, 1.38]	.90	**-0.94 [-1.59, -0.28]**	**-.95**	-0.09 [-1.40, 1.19]	-.50	.05 .11	**-0.97 [-1.58, -0.35]**	**-.97**	-0.88 [-2.06, 0.24]	-.81
Sport/activity	-0.43 [-0.96, 0.11]	-.78	-0.21 [-0.77, 0.34]	-.60	0.69 [-0.36, 1.78]	.76	-.01 -.11	0.13 [-0.52, 0.77]	.53	-0.56 [-1.44, 0.30]	-.75

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Restricted Interest	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
People	**2.83 [0.94, 4.67]**	**.99**	**-2.18 [-3.90, -0.42]**	**-.96**	2.43 [-0.99, 5.76]	.84	.05 .09	**-0.94 [-1.79, -0.10]**	**-.91**	-3.26 [-6.50, 0.07]	-.92

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

As illustrated in Table 4.24, there was a meaningful effect of assessment result on the probability that restricted interests would be noted by diagnosticians and teachers. Similarly, male sex/gender meaningfully predicted concern about restricted interests by both clinicians and teachers. The interaction term for diagnostic observations was large, although a negligible effect could not be excluded with 80% confidence. While the proportion differences appeared similar, the interaction arose from the generally higher proportion of males considered to show impairment here (non-ASD males: 12.1%, ASD males: 41.1%; non-ASD females: 0.0% and ASD females: 26.4%).

Table 4.24

Logistic Regression Predicting Restricted Interests by Assessment Result, Sex/Gender, and Their Interaction: Criterion B3, Diagnostic Observations and Teacher Report

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Source	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Diagnostic observations	**3.14 [1.62, 4.57]**	**1.00**	**2.18 [0.66, 3.60]**	**1.00**	-3.01 [-5.86, -0.06]	-.95	.12 .15	**0.67 [0.19, 1.13]**	**.94**	**3.63 [0.74, 6.45]**	**.98**
Teacher report	**1.19 [0.55, 1.83]**	**.99**	**1.05 [0.40, 1.69]**	**.98**	0.41 [-0.85, 1.69]	.63	.33 .16	**1.25 [0.65, 1.83]**	**.99**	0.83 [-0.33, 1.93]	.81

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

Sensory Behaviour. There was strong evidence (i.e., a negligible difference could be excluded with 80% confidence) that a positive ASD assessment result was associated with parent-reported difficulty in all sensory domains except *Auditory seeking* behaviours (Table 4.25). Compared to parents of males, there was weak evidence that parents of females had a higher probability of reporting *Tactile avoiding, Olfactory avoiding,* and *Visual seeking* behaviours (although a negligible effect could not be excluded with 80% confidence).

Oral avoiding and *Visual avoiding* behaviours were more strongly associated with positive ASD result for males than for females ($RR_M = 3.9$, $RR_F = 1.3$; $RR_M = 8.0$, $RR_F = 0.7$), and there was weaker evidence of this in *Visual seeking* ($RR_M = 2.5$, $RR_F = 1.3$) and *Auditory avoiding* behaviours ($RR_M = 2.5$, $RR_F = 1.6$).[52] Indeed, the sex/gender difference observed in *Visual avoiding* behaviours was mostly driven by the sex/gender difference in non-ASD children (females > males).

[52] A summary of the behaviours in which a meaningful interaction was found between assessment result and sex/gender (i.e., the $HDI_{80\%}$ lay completely outside the ROPE) is presented in Appendix I.

Table 4.25

Logistic Regression Predicting Sensory Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion B4, Parent Report

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Auditory: seeking	0.44 [-0.27, 1.12]	.74	0.03 [-0.68, 0.72]	.45	0.19 [-1.21, 1.55]	.53	.05 .03	0.12 [-0.54, 0.76]	.52	-0.06 [-1.30, 1.18]	-.48
Auditory: avoiding	**1.40 [0.97, 1.82]**	**1.00**	-0.09 [-0.54, 0.33]	-.49	0.86 [-0.00, 1.71]	.87	.43 .23	0.33 [-0.14, 0.78]	.74	-0.52 [-1.24, 0.22]	-.77
Tactile: seeking	**1.46 [0.94, 1.97]**	**1.00**	0.38 [-0.15, 0.89]	.76	-0.52 [-1.56, 0.50]	-.71	.27 .33	0.12 [-0.32, 0.55]	.52	0.64 [-0.32, 1.56]	.78
Tactile: avoiding	**1.20 [0.77, 1.62]**	**1.00**	-0.44 [-0.87, -0.01]	-.84	-0.02 [-0.86, 0.85]	-.45	.18 .29	-0.45 [-0.91, 0.00]	-.84	-0.43 [-1.16, 0.29]	-.72
Olfactory: seeking	**1.29 [0.41, 2.14]**	**.98**	0.17 [-0.70, 1.03]	.55	-0.57 [-2.23, 1.17]	-.64	.08 .11	-0.10 [-0.72, 0.50]	-.50	0.44 [-1.17, 2.02]	.61
Olfactory: avoiding	**1.71 [0.94, 2.45]**	**1.00**	-0.61 [-1.38, 0.12]	-.81	0.53 [-0.96, 2.01]	.65	.19 .17	-0.35 [-0.85, 0.16]	-.74	-0.88 [-2.30, 0.47]	-.77
Oral: seeking	**1.09 [0.62, 1.55]**	**1.00**	-0.34 [-0.81, 0.13]	-.74	-0.33 [-1.27, 0.60]	-.62	.19 .28	-0.50 [-0.94, -0.05]	-.88 *	-0.18 [-1.00, 0.65]	-.55
Oral: avoiding	**1.04 [0.52, 1.54]**	**.99**	-0.33 [-0.84, 0.19]	-.71	**1.35 [0.32, 2.36]**	**.94**	**.29 .07**	0.34 [-0.10, 0.82]	.75	-0.99 [-1.93, -0.10]	-.89 *
Visual: seeking	**0.71 [0.19, 1.25]**	**.94**	-0.61 [-1.14, -0.08]	-.89	0.82 [-0.22, 1.88]	.81	.15 .06	-0.20 [-0.69, 0.29]	-.73	**-1.02 [-1.97, -0.10]**	**-.90 ***

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Visual: avoiding	**0.96 [0.11, 1.77]**	**.93**	**-1.05 [-1.86, -0.21]**	**-.95**	**2.88 [1.21, 4.45]**	**.99**	**.14 -.06**	0.39 [-0.23, 1.01]	.73	**-2.47 [-4.00, -0.99]**	**-.99 ***

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

* Indicates that the sex/gender difference differed substantially between the ASD and non-ASD groups.

As summarised in Table 4.26, there was a higher probability that diagnosticians and teachers reported concern regarding sensory behaviour for children who received ASD diagnoses. Sex/gender did not predict the probability that concern would be raised by either party and there was equivocal evidence of an interaction between sex/gender and assessment result on the probability of concern in either case.

Discussion

The present study examined phenotypic differences between male and female children aged between 2 and 18 years who were referred for and undertook formal developmental assessment and were (a) diagnosed with ASD, or (b) not diagnosed with ASD, despite having many ASD-related characteristics. An important contribution of this study was the consideration of fine-grained diagnostic information from diverse sources including parents, teachers, and the observations of diagnosticians. Both ASD and non-ASD females were least likely to meet Criterion B3 (restricted interests) and less likely than their male counterparts. Furthermore, there was strong evidence that their restricted interests differed to those of males. Non-ASD (subclinical) females, despite showing behaviour that partially met at least one social communication and one RRBI criterion, were found to lack sufficient atypicality for diagnosis in some domains (particularly Criterion A1, social-emotional reciprocity). Some ASD behaviours were more likely to be reported for females compared to males, while others were less likely. Notably, there were some behaviours, diagnostic of ASD and often seen in males, that were not commonly observed in females (e.g., difficulties with imaginative play). Finally, other behaviours were found to be more *diagnostic* (i.e., more strongly associated with assessment result) for males, and a similar pattern was also observed for females. Together, these findings may assist in understanding why females may be less likely to be diagnosed with ASD than males and provide direction for future assessment protocols.

Table 4.26

Logistic Regression Predicting Sensory Behaviour by Assessment Result, Sex/Gender, and Their Interaction: Criterion B4, Diagnostic Observations and Teacher Report

	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)	ASD: M-F		Non-ASD: M-F	
Source	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaningful)	LOR [$HDI_{80\%}$]	P (meaning.)		LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Diagnostic observations	**0.78 [0.27, 1.30]**	**.96**	-0.13 [-0.66, 0.38]	-.53	0.33 [-0.68, 1.38]	.62	.16 .11	0.03 [-0.45, 0.50]	.43	-0.30 [-1.23, 0.60]	-.61
Teacher report	**0.93 [0.24, 1.60]**	**.95**	0.14 [-0.54, 0.81]	.53	0.38 [-0.96, 1.76]	.60	.19 .12	0.32 [-0.28, 0.95]	.68	-0.06 [-1.24, 1.18]	-.48

Note. Positive LORs (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LORs (sex/gender) = greater probability of being reported for males. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside the ROPE. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction. Prop. Diff. = proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). LOR = log odds ratio.

Sex/Gender Differences in ASD Presentation

Consistent with the theoretical framework outlined by Lai et al. (2015), evidence of sex/gender differences in the present study was most apparent at the level of specific behaviours. Parents of females were meaningfully more likely to report difficulty in some behaviours (e.g., difficulties with eye contact) and in specific stereotypical mannerisms and interests. Surprisingly, parents of females were more likely to raise concern in several behaviours within nonverbal behaviour (A2) than parents of males, but despite this, these behaviours did not translate to a higher probability that females would meet this criterion during the diagnostic assessment. This is inconsistent with Hiller et al. (2014) who reported that males had greater or at least equally severe difficulty concerning nonverbal behaviour. However, the findings of the present study are broadly consistent with those of Bitsika and Sharpley (2019), who also found that numerous social difficulties were more severe among females. Unlike the present study, Bitsika and Sharpley's findings may be the result of matching males and females based on ADOS-2 scores (i.e., a measure of observable ASD behaviours, thought to be less severe among females and thus including only females with greater ASD difficulty). In the present study, this discrepancy may be attributable to the inclusion of subclinical (non-ASD) females. Specifically, in some areas (e.g., understanding of nonverbal behaviour), overall sex/gender differences were driven by sex/gender differences in the non-ASD children. Differences in other areas (e.g., facial expression) may be the result of the bottom-up approach (i.e., reliance upon more loosely operationalised difficulties, rather than very specific examples provided in questionnaires, as in Study 1). Alternatively, there may be different 'requirements' for referral for a male versus a female, and females may need to display a greater number of difficulties than males in the social domains in order for referral to be made (Dworzynski et al., 2012).

Parents of females were also more likely than parents of males to report several sensory avoiding behaviours. Findings surrounding sensory sex/gender differences have been inconsistent, both broadly (Frazier & Hardan, 2017; Lai et al., 2011) and in specific sensory sensitivities (Bitsika et al., 2018). As females were found to be slightly more likely to meet Criterion B4 in the present study, further investigation into whether elevated sensory difficulties constitute an important element of the female presentation is warranted.

Consistent with previous findings (e.g., Beggiato et al., 2017; McFayden et al., 2019), females were less likely than males to meet Criterion B3 and demonstrate restricted interests. As also reported by Hiller et al. (2014) and McFayden et al. (2018), sex/gender differences were present in the nature of restricted interests, with males more likely to present with obsessions with vehicles and screens, and females more likely to be interested in other people and craft activities. However, a novel finding of the present study is that some interests were more strongly associated with ASD diagnosis for males (e.g., random objects or people) or females (e.g., animals or specific programs/characters). As previously suggested (e.g., Nowell et al., 2019), restricted interests that are less overtly atypical in orientation may be diagnostically overlooked.

In addition to presenting with different restricted interests, males and females also presented with different parent-reported stereotypical behaviours. In contrast with other findings (e.g., Study 1 and Antezana et al., 2018), this was least apparent in stereotypical body use behaviours, with the only meaningful difference being a higher probability of repetitive body use mannerisms (such as hair twirling or nail biting) among females. This result is somewhat consistent with the findings from the only existing study on sex/gender differences in fine-grained stereotypical behaviours, conducted by Antezana et al. These authors reported a significantly higher incidence of *Pull[ing] hair/skin* ($d = 0.34$) and

Rubb[ing]/scratch[ing] self (d = 0.20) among females.[53] Therefore, while it was unexpected that, in the present study, the best estimate of sex/gender difference for self-injurious behaviour was within the negligible range (d = 0.04, $HDI_{80\%}$ = [-0.53, 0.62]), this may be the result of the conceptual overlap between some self-injurious behaviours and repetitive body use. While these behaviours were considered separately in the present study, they overlapped in the items examined by Antezana et al. Concerning other body use behaviours, these authors found that hand/finger mannerisms were weakly more prevalent among males (d = 0.24), whereas no meaningful sex/gender difference was found in this area in the present study. Further, the best estimate of sex/gender difference in parent-reported stereotypical object use in the present study was in the negligible range (d = 0.07, $HDI_{80\%}$ = [-0.57, 0.69]), and weak evidence suggested greater a greater probability of females presenting with grouping (d = -0.45, $HDI_{80\%}$ = [-0.94, 0.06]), and repetitive play behaviours (d = -0.61, $HDI_{80\%}$ = [-1.28, 0.08]). These findings contrast with the results of Study 1 and other investigations (e.g., Antezana et al., 2018; Reinhardt et al., 2015) which have suggested that stereotypical object use may be more pronounced among males. It is likely that (a) the examination of specific object use behaviours rather than the subdomain level, (b) differences in operational definitions applied and whether they were restricted by diagnostic instruments, and (c) the inclusion of non-ASD children, may account for the differences in findings.

Finally, males had a higher probability of presenting with parent-reported speech mannerisms in general, and repetitive speech and echolalia specifically, while females were more likely to present with neologisms, accents, and talking to oneself. Sex/gender differences in speech and language use have remained largely overlooked in the literature,

[53] To compare my findings with those of Antezana et al. (2018), I calculated an effect size (Cohen's f) from their reported ANOVA results. As f is on the scale of half Cohen's d, I doubled those effect sizes for comparison with my results.

but there is some recent evidence to suggest that there may be differences in linguistic features associated with storytelling (see Boorse et al., 2019). To my knowledge, the present findings regarding speech stereotypy profiles are novel. While it is unknown why stereotypical speech profiles may differ by sex/gender, this may result from differences in language development, or a consequence of camouflaging and imitation (e.g., accents may be imitated from others) and is worthy of further exploration.

Females Under the Diagnostic Threshold

In addition to presenting with different stereotypic behaviours and restricted interests to males, females were less likely to meet Criteria B1 and B3 than their male counterparts, regardless of diagnostic result. In some instances, the stereotypical behaviours that were more common in females were less strongly associated with ASD diagnosis than those typically presented by males. Indeed, repetitive body use and neologisms were more common among non-ASD females than any other group, and hand-mannerisms, rocking/jumping, and repetitive play were as common in non-ASD females as children with ASD, irrespective of sex/gender. This suggests that sex/gender specific profiles of stereotypical behaviours may exist, and that those presented by females may not perfectly coincide with ASD conceptualisation (see also Lai, Lombardo, et al., 2015; van Wijngaarden-Cremers et al., 2014). It is possible, therefore, that these sex/gender specific profiles may contribute to under-detection of ASD in females.

As expected, non-ASD females lacked a number of difficulties that were present among ASD females. In particular, non-ASD females lacked sufficient difficulty in social-emotional reciprocity (Criterion A1; where non-ASD males were approximately half as likely to partially meet and twice as likely to fully meet the criterion as non-ASD females). Two possible reasons for this can be deduced from this study: first, a number of parent-reported A1 difficulties were more strongly associated with diagnosis for males, and second, males

were more likely to display a number of difficulties in this area during assessment. Both of these issues are discussed in greater depth below. Importantly, evidence suggested that non-ASD females were more likely to have parent-reported concern than non-ASD males in the social communication and RRBI domain, whereas for diagnostician observation and teacher report, non-ASD males were weakly more likely to raise concern. Together, these findings suggest that, for sub-clinical females with autistic traits, insufficient evidence of difficulty observed during assessment and/or in the school environment may contribute to why an ASD diagnosis is not be made.

Discrepancies Between Environments and Report Sources

Consistent with the notion that the presentations of autistic females may differ according to their environment, levels of concern differed according to the source of reporting. Unlike in parent report (where ASD females had a similar probability of reported concern compared to ASD males across both domains), teachers were more likely to raise concern for males. There was also weak evidence for this for diagnostician observations. Sex/gender predicted the presence of some specific behaviours according to diagnostic observations and teacher report, which showed similar trends. Unlike parent report in which females were more likely than males to have several difficulties reported, there was only one diagnostician observed behaviour in which females were more likely to raise concern: atypicality in facial expression. For the remaining behaviours, there was either no meaningful sex/gender difference, or males were more likely to raise diagnosticians' concern. To a greater extent than what was seen in diagnostician observation, teacher concern was meaningfully higher if the child was male. Similarly, Hiller et al. (2014) identified no specific domains in which teachers reported more concern for girls than for boys. Other studies have also shown that teachers may report significantly less concern for girls than boys, and that this sex/gender discrepancy is larger than in parent reporting (Mandy et al., 2011; Posserud et

al., 2006). Although the present investigation did not examine concern regarding externalising behaviour, it is possible that, if indeed the behaviour of males with ASD traits is more disruptive in the school environment (Hiller et al., 2014; Mandy et al., 2011), this may result in their difficulties becoming more salient to teachers and hence result in increased reporting.

A smaller number of meaningful sex/gender differences were observed according to diagnosticians' observation compared to teacher-reported behaviours. Indeed, for many behaviours, sex/gender did not meaningfully predict the probability that they would be observed by the diagnostician. However, males were more likely than females to display several specific difficulties within social-emotional reciprocity, stereotypical object use and restricted interests during assessment. Therefore, diagnostic instruments which rely on assessment observations may underestimate ASD features among females who do not present with sufficient observable and atypical behaviours during the assessment period. This concern has also been raised by other authors who have found that observation schedules, such as the ADOS-2, may lack sensitivity to the female ASD presentation because of differences in outward presentation (Adamou et al., 2018; Lai et al., 2011).

A substantial number of behaviours were more likely to be reported by parents of females than males, whereas this was not the case for diagnostic observations or teacher report. Although these data should be treated with caution due to differences in the thoroughness of reporting and how the information was collected across sources, they raise questions as to how such data should be obtained and from whom. The discrepancies across reporting sources suggest that there may be differences in a female's ASD presentation according to her environment and lends support to the camouflage hypothesis; i.e., females may be more highly motivated or better able to conceal and compensate for ASD related difficulties in some environments, such as at school and during assessment (Hull et al., 2017;

Rynkiewicz et al., 2016). Thus, parents may see the *true* ASD presentation (i.e., the genuine and undisguised difficulties) or emotional consequences of camouflaging, whereas at school, females may display fewer ASD-related behaviours and therefore raise little concern by teachers.

It is important to note that a number of factors may influence whether a behaviour is reported during the assessment process, and therefore, if a behaviour was not reported, we cannot exclude the possibility that it was present. It is understood that gender expectations may affect how a behaviour is interpreted when it occurs (Kreiser & White, 2014); for example, spinning or finger twinkling mannerisms or a desire for deep-pressure hugs may be interpreted as endearing for a young girl, but atypical for a young boy. Thus, if these behaviours are questioned during assessment, parents of girls may normalise the behaviours in their response (e.g., "no more than any other child"). In addition to differences in how a behaviour is interpreted, the sensitivity of the caregiver to the behaviour and its disruptiveness will affect whether it is reported. Furthermore, ASD features may differ in manifestation, which may not be adequately captured by all diagnostic interviews (e.g., while a girl may know that she needs to offer comfort when someone is hurt, she may feel very uncomfortable doing so). Finally, the genuine absence of a particular ASD difficulty at the time of assessment does not preclude it from coming to light at later stages (as seen in Study 2a). The findings of the present study provide an important impetus for future research targeting these issues.

Sex/gender Differences in Behaviours Associated with Diagnosis

Some behaviours were less strongly associated with assessment result than others reported from the same source. This is not necessarily surprising, as some behaviours, such as *Possessiveness/difficulty losing* may be considered less developmentally atypical or less strongly associated with ASD than other behaviours, such as difficulty *Adjusting [one's]*

behaviour for the situation. However, unexpectedly, many behaviours were more *diagnostic* for males or females – that is, the presence of particular behaviours was more strongly predictive of a positive ASD diagnostic result for males than females or vice versa. There was unequivocal evidence of an interaction for 18 behaviours, 12 of which showed a larger difference in the probability of the behaviour between ASD and non-ASD children for males than females (refer to Appendix I for a summary of meaningful interactions). A minority of these (22.2%) were found in the social communication domain, with the majority in the RRBI domain and within stereotypic behaviours and restricted interests in particular. The presence of interactions in the RRBI domain is consistent with the findings of Duvekot et al. (2016), wherein RRBI difficulties were more strongly associated with ASD diagnosis for males than females. The present study builds on this finding, identifying specific behaviours across all domains which may contribute unevenly to ASD diagnosis for males and females.

While interactions were scattered across all sources of information, it is the diagnostician who must draw on all reported information to decide upon the assessment result. Therefore, it is possible that (a) diagnosticians have a different view of which behaviours are most diagnostic for males and females, suggesting they have an implicit understanding of differing ASD presentations, and/or (b) diagnosticians interpret behaviours differently depending on the child's sex/gender. The latter is not necessarily evidence of bias, as this could also reflect differences in the typical developmental trajectories of males and females. Future research should explore these possibilities in the context of ASD diagnosis.

Strengths and Limitations

A particular strength of this study was the bottom-up approach to data collection. This enabled movement beyond existing psychometric instruments which may perpetuate traditional views of ASD presentation and lack sensitivity to the difficulties of females (see Hiller et al., 2014; Lombardo et al., 2019). Another strength was the inclusion of children

under the diagnostic threshold (i.e., who were formally assessed and received a negative ASD result) which allowed for the investigation of the sex/gender differences in the ASD features that are commonly absent and therefore reasons that ASD diagnoses are not made. Examination of a large array of ASD behaviours as reported by several different sources allowed for detailed study of sex/gender differences in ASD presentation.

However, several limitations restrict the conclusions that can be made from the current study. First, sample sizes were small (particularly for children diagnosed with ASD only at a second assessment). Second, data were collected from a single private clinic in Adelaide and from the assessment reports of one group of diagnosticians (n = 7). Thus, the extent to which these results are generalisable is unknown. Second, while a primary contribution of this study was the investigation of the presentations of females who have many ASD traits but are not diagnosed, it only included children who presented for specialist assessment. As a result, a significant subgroup of children with many ASD traits, or who may meet diagnostic criteria if they came to clinical attention, were not included. This is likely to be a disproportionate issue for females. By virtue of the fact that, for these individuals, their ASD traits have been overlooked, normalised, or otherwise explained, it may be challenging for research to identify and explore their experiences. Finally, given many behaviours were coded as a binary variable (i.e., significant concern reported or not), it may be of interest to examine the degree to which each difficulty presents for males and females, as well as the functional limitations that difficulty poses for the individual, either in isolation or in combination with other difficulties. Exploring these differences may help to further clarify the female ASD presentation and the clinical significance of difficulties of females under the diagnostic threshold.

Conclusions and Future Directions

In this study, a large number of sex/gender differences were found in the probability of presenting with particular autistic behaviours, as reported by parents and teachers and observed by diagnosticians during assessment. For females who met ASD criteria at a second assessment after an initial negative result, increased difficulties in (a) conversation content, and (b) resistance to change, routines and rituals, most meaningfully contributed to ultimately meeting the criteria.

For children who attended one ASD assessment, numerous sex/gender differences emerged in whether specific behaviours were reported, and importantly, by whom. Females were substantially less likely to meet Criterion B3 (restricted interests), and this was especially the case for non-ASD (subclinical) females. There was evidence that the nature of restricted interests and stereotypical behaviour profiles differed by sex/gender, and many of these behaviours were less strongly associated with ASD diagnosis than those presented by males. Non-ASD females also lacked a number of social-emotional reciprocity difficulties. Additionally, there was a meaningful contrast in the levels of concern of parents and the observations of diagnosticians and teachers for females, suggesting that camouflage in social environments may contribute to a negative ASD result, or at least require greater impairment in other areas in order to qualify for diagnosis. Finally, for a number of behaviours, their presence was more indicative of a positive diagnosis for males than females or vice versa, suggesting gender differences in how atypical behaviours are perceived. Given these results, it may be necessary to frame diagnostic criteria more flexibly so as not to exclude females who do not present with certain difficulties, or for whom less concern is raised in social environments.

This study has illustrated that certain ASD-related behaviours may be interpreted differently by diagnosticians depending on the sex/gender of the child, and thus these

behaviours may contribute to diagnostic outcome to different extents for males and females. Little is currently known about sex/gender differences in diagnosticians' interpretation of ASD-related behaviours, and any sex/gender-related expectancy bias that may exist. Further, diagnostic decision-making regarding females presenting query ASD and, more broadly, the challenges that may arise within this process, remain largely unknown. Exploring these issues will therefore be the focus of Study 3.

Chapter 5: Study 3

A Mixed-Methods Investigation of Diagnostician Sex/Gender-Bias and Challenges in Assessing Females for Autism Spectrum Disorder

Overview

Growing evidence suggests that ASD is underdiagnosed among females, rendering some individuals unable to access the specialised support they may require (Hull et al., 2020). The extent of this issue is highlighted in the discrepancy between population prevalence estimates (three males for every female) and in individuals who have received a diagnosis (four to five males for every female; Loomes et al., 2017). Females may need to display greater difficulty in ASD symptomology, adaptive behaviour, and/or intellectual ability compared to males in order to receive an ASD diagnosis (Duvekot et al., 2016; Dworzynski et al., 2012; Ratto et al., 2018; Russell et al., 2011).

Despite the importance of clinical judgement in ASD assessment and diagnosis, little is currently known about challenges faced by diagnosticians when the client is female (Attwood et al., 2006). Clinical judgement in discerning the presence or absence of ASD relies on the interpretation of ASD behaviours in the context of typical development and the disorder conceptualisation, and some authors have suggested that the sex/gender of a child may result in biases herein (Lai & Szatmari, 2019). However, research investigating sex/gender bias during assessment, diagnosticians' perceptions of phenotypic sex/gender differences, challenges inherent in the assessment of female clients for ASD and how these may be addressed, is limited. Exploring these issues will likely broaden understanding of the underdiagnosis of females and inform future research and training of diagnosticians. The current study will therefore examine sex/gender expectancy bias and diagnosticians' perceptions of challenges in assessing females for ASD.

It is now understood that ASD is more prevalent among females than once thought, and autistic females may be less likely to meet criteria than males if formal assessment is pursued (Wilson et al., 2016). The reasons for the underdiagnosis of this condition in females have received considerable empirical attention in recent years and there is growing consensus that underdiagnosis may be partly the result of subtle sex/gender differences in ASD presentation. In particular, females may present with less overtly atypical restricted interests, fewer stereotyped motor behaviours, greater social motivation, and the use of compensation and masking strategies (i.e., camouflaging; Hull & Mandy, 2017). Characteristics of 'classic' ASD (i.e., the typical perception of the ASD presentation) may include behaviours that are more disruptive or obviously atypical, leading to earlier referral for assessment and more timely diagnosis (Hiller et al., 2014).

Consistent with the above, recent evidence suggests that clinicians may conceptualise ASD slightly differently for males and females. That is, while clinicians have reported that girls and boys present with similar levels of difficulty in core ASD symptoms (i.e., social communication and repetitive behaviour), sex/gender differences may be present in secondary or associated features (e.g., emotional regulation, social motivation) and management of ASD difficulties (e.g., internalising or externalising behaviours; Muggleton, et al., 2019). Clinicians have also noted several sex/gender differences in the manifestation of repetitive and restricted behaviours in particular, such as the focus of restricted interests (Jamison et al., 2018).

If, as evidence suggests, there are qualitative differences between the presentation of 'classic' male ASD and that of many females, this may pose challenges for diagnostic assessment. Of particular concern, assessment instruments based on the androcentric literature may not reliably detect the female presentation and thus misrepresent symptom severity or overlook less common expressions of symptoms (Lai, Lombardo, et al., 2015).

However, the vast majority of investigations into the reasons for underdiagnosis of ASD in females have relied upon these instruments, therefore including only diagnosed females who present most similarly to the classic ASD presentation (Bargiela et al., 2016). Using current assessment instruments to examine underdiagnosis is therefore unlikely to capture (a) the specific challenges faced by females with ASD, and (b) the difficulties encountered by diagnosticians in identifying these challenges or presentations. Given the majority of these instruments has been designed with the male presentation in mind, the onus rests on the diagnostician to interpret and translate their results so they may be understood within the context of the female ASD presentation. It remains unclear to what extent diagnosticians rely on and/or adapt these instruments for use with female clients.

In addition to any bias in assessment instruments, biases in the interpretation of ASD-related behaviours among clinicians, parents, or other referring parties, may also contribute to the under-detection of ASD in females (e.g., Kreiser & White, 2014). These may result from sociocultural expectations of how males and females typically behave at any given developmental stage.

Two recent studies have experimentally examined sex/gender bias in the perceptions of ASD behaviours among potential referrers: members of the general population (Geelhand et al., 2019) and educators (Whitlock et al., 2020). In the former study, participants considered descriptions of autistic behaviours to predict future abnormality more strongly if the description referred to a boy rather than a girl. However, results showed that sex/gender did not meaningfully affect the degree of concern raised about these behaviours. In contrast, Whitlock et al. (2020) found that educators presented with vignettes (including a typical ‘male’ and ‘female’ ASD presentation in which the sex/gender of the child described in the vignette was randomly assigned) were less likely to identify children presenting with the female ASD presentation as being autistic. Significant evidence of bias against detecting

ASD for females was only detected in the female phenotype and not in the male phenotype vignette.

Despite this evidence of bias in whether concern is raised, little is known about any bias among diagnosticians at the time of ASD assessment and, if it exists, its role in whether an ASD diagnosis is provided. This gap in the literature remains despite reports from autistic women and adolescent girls who believe they faced additional difficulties during the assessment process due to diagnosticians' lack of familiarity with how females may express their ASD difficulties (Baldwin & Costley, 2016; Bargiela et al., 2016; Navot et al., 2017).

Sex/gender-related clinician expectancy bias, or a tendency to consider and/or diagnose a condition when it is more common in a particular sex/gender or has features more closely aligned with normative behaviours of this sex/gender (Hartung & Widiger, 1998) has been documented in a number of psychiatric conditions (Potts et al., 1991; Worell & Robinson, 2009). For ASD, such bias may manifest in two ways: (a) the degree to which clinicians are inclined to consider ASD, and (b) in the interpretation of ASD-related behaviours.

Given that ASD is historically more prevalent amongst males, clinicians may be primed to consider ASD more frequently for males, thus perpetuating the higher diagnostic rates. Theories such as the Extreme Male Brain theory may contribute to the perception that ASD is primarily a 'male' condition, in conceptualising ASD as an extreme version of male (systemising) neurodevelopment and arguing that males are therefore more vulnerable to developing ASD (Baron-Cohen, 2002; Baron-Cohen & Hammer, 1997). In support of this theory, ASD traits have been found to be distributed unevenly between the sexes in typically developing children, with boys generally demonstrating higher levels (Constantino & Todd, 2003).

Additionally, sex/gender expectancy biases may mean that unusual behaviours are more likely to be interpreted as indicative of ASD for males compared to females. For example, social avoidance may be interpreted as 'shyness', 'anxiety', or 'rudeness' for girls, but 'aloofness' or 'social disinterest' in boys. The latter interpretation is likely to be more strongly suggestive of developmental abnormality and possibly ASD. In support of this, a small number of investigations has provided evidence that behaviours may be interpreted differently depending on the sex/gender of the child. Russell et al. (2011) examined female sex/gender as one of many possible barriers to timely diagnosis of ASD in their retrospective analysis of a longitudinal cohort study, and found that males are more likely to be diagnosed with ASD compared to females with equally severe symptoms. Similarly, clinicians reviewing educational records in a population cohort were less likely to classify girls as having ASD than boys (Giarelli et al., 2010).

Under-detection and thus underdiagnosis of ASD in females may be the result of a lack of community awareness and thus whether difficulties are identified and referral for specialist assessment made (Hull & Mandy, 2017; Whitlock et al., 2020). However, of interest to the present study are diagnosticians' views of the obstacles to accurate diagnosis and how these are managed, once a referral has occurred. To date, no studies have adopted an experimental methodology to examine sex/gender-related diagnostician biases or decision-making in the context of ASD. Similarly, in exploring the challenges in ASD diagnosis for females, the voices of diagnosticians have been largely unheard (with two recent exceptions: Jamison et al., 2018; Muggleton et al., 2019). In this study, I adopted both an experimental and mixed methods design to address the following research questions:

1. Is there a sex/gender-bias in the likelihood that diagnosticians will consider ASD criteria met and their impressions of the severity of the presenting difficulties? How does this relate to the diagnosis made?

2. Do diagnosticians believe there are sex/gender differences in the typical presentation of ASD? If so, what are these differences?
3. How do diagnosticians rate their familiarity with sex/gender differences in ASD presentation and how confident are they in assessing females for ASD?
4. What do diagnosticians perceive as challenges in assessing females for ASD?
5. If challenges are perceived, how have diagnosticians changed their assessment practice to circumvent these?

Method

Participants

Only clinicians specifically trained in ASD diagnosis were recruited for participation in this study. Recruitment was conducted in two waves. In the first wave, delegates who registered to attend a conference workshop presented by R.Y. and J.T. about the presentation of ASD in females were recruited. Fourteen delegates completed the questionnaire prior to the workshop and consented to having their responses used for research. It is unknown how many delegates were eligible to participate and therefore the response rate is also unknown. In the second wave, Australian and New Zealander ASD diagnosticians (n = 33) were invited to participate via email through local autism associations and private clinics (n = 44 organisations), and via advertisement on social media.[54] The rate of recruitment in this wave is also unknown. Eligible participants were financially reimbursed for their time.

Characteristics of participants (hereafter referred to as *diagnosticians*; total N = 47) are presented in Table 5.1. Diagnosticians were predominately female (87.2%) and work as either psychologists (72.3%) or speech pathologists (17.0%). Approximately half of diagnosticians (51.1%) reported working primarily in the eastern states of Australia and a third in South Australia (34.0%). Diagnosticians had an average of 11.48 years of experience

[54] The recruitment flyer is presented in Appendix J.

working with people with ASD (*SD* = 11.00) and 7.24 years of experience conducting assessments for ASD (*SD* = 8.06). Information regarding the number of individuals seen for ASD assessments per age group in an average six-month period was available for 74.5% of diagnosticians. Of these, all reported conducting between one and three assessments for school age children (age 5-11 years), 74.3% for children in early childhood (age 0 to 4 years), 71.4% for adolescents (age 12-17 years) and 59.6% for adults.

Table 5.1

Characteristics of Diagnostician Participants (N = 47)

Sex/gender	Female 87.2% (*n* = 41)	Male 12.8% (*n* = 6)		
Profession	Psychologist 72.3% (*n* = 34)	Speech path. 17.0% (*n* = 8)	Paediatrician 4.3% (*n* = 2)	
	Occ. therapist 4.3% (*n* = 2)	Psychiatrist 2.1% (*n* = 1)		
State	South Aust. 34.0% (*n* = 16)	Victoria 23.4% (*n* = 11)	Queensland 17.0% (*n* = 8)	New South Wales 10.6% (*n* = 5)
	Western Aust. 8.5% (*n* = 4)	Tasmania 2.1% (*n* = 1)	New Zealand 4.3% (*n* = 2)	
ASD experience (years)	≤ 2 12.8% (*n* = 6)	3-5 27.7% (*n* = 13)	6-10 21.3% (*n* = 10)	
	11-20 23.4% (*n* = 11)	21-30 6.4% (*n* = 3)	> 30 8.5% (*n* = 4)	
ASD assessment experience (years)	≤ 2 27.7% (*n* = 13)	3-5 34.0% (*n* = 16)	6-10 17.0% (*n* = 8)	
	11-20 10.6% (*n* = 5)	21-30 8.5% (*n* = 4)	> 30 2.1% (*n* = 1)	

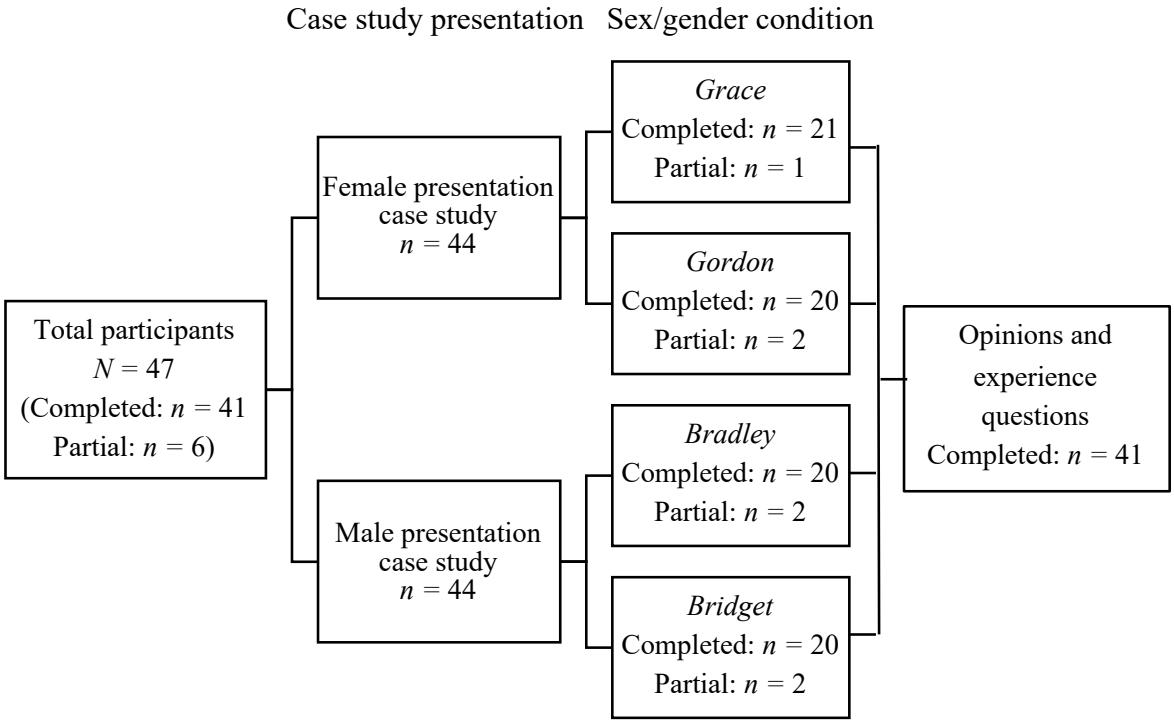
Materials

Questionnaire

The online questionnaire was developed using *Qualtrics* software and consisted of three sections (refer to Figure 5.1 for a summary of the questionnaire structure and the number of diagnosticians who completed each component).[55] Items in the first section assessed diagnostician demographics and experience in ASD assessment. The second section comprised two extended case studies which were constructed based on a large number of clinical cases. One case study was designed to reflect a mild 'male' presentation, and the other a mild 'female' presentation. Previous empirical findings surrounding sex/gender

Figure 5.1

Summary of Questionnaire Structure



[55] The questionnaire is presented in Appendix K.

differences in ASD presentation (including those outlined in Chapter 1 and identified in Chapters 3 and 4) and clinical, qualitative, and autobiographical works (e.g., Attwood et al., 2006; Holliday-Willey, 2015; Kanfiszer et al., 2017) were used to inform the construction of these case studies. Key elements of the female case study included internalising behaviours and anxiety (Oswald et al., 2015), compulsive and self-injurious behaviours (Antezana et al., 2018), superficial skills in conversation reciprocity and friendship formation (Hiller et al., 2014), gender appropriate restricted interests (Nowell et al., 2019), and imagination in play (Knickmeyer et al., 2008), etcetera.

The sex/gender (and pseudonym) of the child presented in the case studies was randomly assigned, such that diagnosticians answered questions about *Bradley* or *Bridget* for the male presentation case study and *Grace* or *Gordon* for the female presentation case study. Each diagnostician was presented with a pair of cases (i.e., one male presentation and one female presentation) which were otherwise identical across conditions.[56] For both case studies, information relevant to each ASD criterion outlined in the *Diagnostic and Statistical Manual for Mental Disorders - 5th Edition* (DSM-5; American Psychiatric Association, 2013) was provided and diagnosticians asked to indicate whether each criterion was met, not met, or partially met (i.e., demonstrating some difficulty but not enough to fully meet the criterion). Diagnosticians were also asked to rate their confidence that each criterion was met from 0 (*not at all confident*) to 100 (*extremely confident*), the severity of the child's difficulty within each criterion from 0 (*very mild*) to 100 (*extremely severe*) and the level of support required in the social-communicative and RRBI domains (Level 1-3 as defined by the DSM-5; American Psychiatric Association, 2013). Follow-up questions about the most appropriate

[56] Only the child's pseudonym, gendered pronouns (i.e., he, him, his/she, her) and names of friends (matched with the pseudonym sex/gender) were altered across conditions. In the construction of the case studies, interests were selected to be as gender neutral as possible (e.g., drawing and animals for the female presentation case study and *Lego* and reading for the male presentation case study).

diagnosis/es for the child were also posed. To guard against ceiling and floor effects, the information provided was deliberately ambiguous as to whether any given criterion was met. A number of behaviours not traditionally associated with ASD (e.g., repetitive lying) were also included as distractors. No standardised information (e.g., results of cognitive assessments or ASD screening tools) was included. The severity of the child's behaviours was varied across criteria to ensure heterogeneity in responding between criteria and criteria were presented in a random order. Diagnosticians were not able to return to previous questions, so that their later impressions of the child's presentation did not affect their earlier responses. Diagnosticians were excluded if they did not complete at least one case study (n = 2). This left a total of 47 diagnosticians, 41 of whom completed the entire questionnaire.

It was not possible to match ASD severity across the male and female presentation case studies (due to the possibility that ASD psychometric instruments may under-rate the severity of a female presentation and the fictitious nature of the case studies), and therefore no comparisons can be made between them (within-participant variable). Only comparisons between allocated sexes/genders (conditions or between-participant variable) within each case study are presented.

The final component of the questionnaire explored diagnosticians' opinions, experience and confidence in assessing females for ASD using a mixture of closed and open-ended questions. In particular, diagnosticians were asked to rate the extent of their agreement with various statements about sex/gender differences in ASD presentation, constructed based on the findings of previous research (e.g., *Compared to boys, girls are better able to camouflage their difficulties*). This was done to investigate the degree of consistency between diagnosticians' experiences and the findings.

Procedure

The questionnaire was first piloted on a sample of seven ASD diagnosticians to clarify the intelligibility of questions and to guard against ceiling and floor effects within the case study items. Minor adjustments to the case studies and follow-up questions were made following piloting. Specifically, for a small number of criteria, pilot participants were highly confident that the criterion was met and therefore some ASD features were removed to render the conclusion less clear. Although the precise purpose of the study was obscured from all potential participants during recruitment (i.e., the examination of sex/gender expectancy bias in diagnosis), all potential participants were informed that the study was concerned with their experience of assessing females with ASD and were thus equally primed to its broad interests. It was therefore not expected that the responses of delegate diagnosticians would meaningfully differ from those of non-delegate diagnosticians. All participants provided consent for their responses to be used in research and debriefed about the study aims at the conclusion of the questionnaire.

Data Analysis

Case Study Experiment

Hierarchical Bayesian regressions were used for the major analyses and Bayesian logistic regression to interpret the probability of alternative diagnoses being offered and support levels assigned. To maintain consistency across analyses and for ease of interpretation, the effect size of the difference in posterior means is used as the major descriptive statistic and the basis for inference. For differences between two means, the effect size statistic was Cohen's *d* (0.2 indicates a small effect, 0.5 indicates a medium effect, and > 0.8 indicates a strong effect; Cohen, 1988). For differences between two proportions, log odds ratios (LORs) are reported. The region of practical equivalence (ROPE) was defined as ± 0.1 for both *d* and LORs (Kruschke, 2018). For all analyses, male sex/gender was assigned

the positive direction (i.e., M > F = positive), and female sex/gender was assigned the negative direction (i.e., F > M = negative). Therefore, the sign of the effect size indicates the direction of difference (positive for males, negative for females).[57]

Experiences Assessing Females for ASD

Several questions were posed to diagnosticians to explore their perceptions of sex/gender differences in ASD presentation, perspectives of the relative difficulty of assessing males and females, confidence assessing males and females of different ages, and familiarity with the female presentation of ASD. Descriptive statistics, Bayesian *t*-tests and Bayesian Pearson correlation tests were used to analyse these data.

Qualitative Responses

Content analysis was applied to qualitative responses to four open-ended questions: "In your opinion, …

1. … what are three features of ASD that may present differently in females?"
2. … what are three reasons that females with ASD may be underdiagnosed?"
3. … what are the most challenging aspects of assessing a female for ASD?" and
4. "What have you changed about your assessment procedures to circumvent such challenges?"

Responses to these questions were analysed separately using a content analysis approach. For each question, two independent raters (J.T. and R.Y.) separately read all data to establish familiarity and then highlighted particular concepts. Codes were constructed to reflect concepts and refined to produce provisional categories (Hsieh & Shannon, 2005). These were then arranged according to their relationship with other categories to establish final categories and subcategories. Any disagreement was resolved through discussion to

[57] The mathematical formulations underlying the models are presented in Appendix C.

consensus (Braun & Clarke, 2013) and minor refinements were made prior to final coding. Both raters agreed upon the final category structures.

Diagnosticians' responses and therefore final category structures were similar for Questions 2 and 3. Therefore, these structures were combined and presented together. For Question 1, coding was specifically guided by the DSM-5 criteria for ASD. Reponses that did not easily fit into the framework defined by the DSM-5 criteria (e.g., energy level), were coded in separate categories.

The quantitative and qualitative results of this study are presented together to address the research questions. It should be noted that the qualitative results are exploratory and should therefore be viewed as supplementary and secondary to the quantitative results.

Results

Case Study Experiment: Sex/Gender Expectancy Bias

Diagnosticians each viewed two case studies: one designed to reflect the presentation of many females (hereafter referred to as the female presentation case study), and one of many males (male presentation case study). The pseudonym (indicating the sex/gender of the child) was randomly assigned (hereafter referred to as sex/gender condition) to explore whether sex/gender influenced diagnosticians' clinical impressions within each case study. Specifically, diagnosticians rated the probability that each criterion was met, their confidence that each criterion was met and the severity of the child's difficulties in each criterion area.

Endorsement of ASD Criteria

Hierarchical Bayesian regressions[58] were used to examine whether the sex/gender condition (determined by pseudonym) in each case study predicted each of three outcome

[58] The precise nature of the regression differed depending on the outcome variable. The probability that a criterion was met (categorical) was modelled using conditional logistic regression with the relevant probabilities estimated from a linear combination of the predictors via a logit link. The

measures: (a) the probability that a criterion was considered to be met, (b) diagnosticians' confidence that a criterion was met, and (c) diagnosticians' impression of the severity of difficulties within each criterion area. Sex/gender condition, case study presentation (i.e., male or female ASD presentation), and their interaction were defined as predictor variables. Given the research questions, sex/gender condition was of greatest interest in this study. Data for all criteria were added to the model simultaneously with criterion as a random effect (i.e., all predictors, including the intercept, were allowed to vary by criterion). Thus, the analysis estimated both the overall pattern, averaged across criteria, as well as the patterns emerging for each criterion. [59]

Probability that Criteria Were Considered Met. The results of the hierarchical Bayesian regression for the probability that criteria were considered met are presented in Figure 5.2. Across both the male and female presentation case studies, there was weak evidence that diagnosticians were, in general, more likely to deem criteria met in the female sex/gender conditions (i.e., Grace or Bridget; LOR = -0.26, $HDI_{80\%}$ = [-0.45, 0.01]). We can have 79.8% confidence in an effect in this direction (i.e., $P_{(meaningful)}$ = 79.8%) and there is an 18.8% probability that the difference was negligible in size (i.e., $P_{(within\ ROPE)}$ = 18.8%). As depicted in Figure 5.2, the effect of sex/gender condition on the probability that any specific criterion was considered met was largest in Criteria A2, B1, and D in the male presentation case study, and in B3 for the female presentation case study.

There was equivocal evidence as to whether the probability of deeming criteria met differed between the male and female presentation case studies (LOR = -0.03, $HDI_{80\%}$ = [-0.30, 0.23], $P_{(within\ ROPE)}$ = 44.2%; an effect in either direction could not be excluded with 80%

---

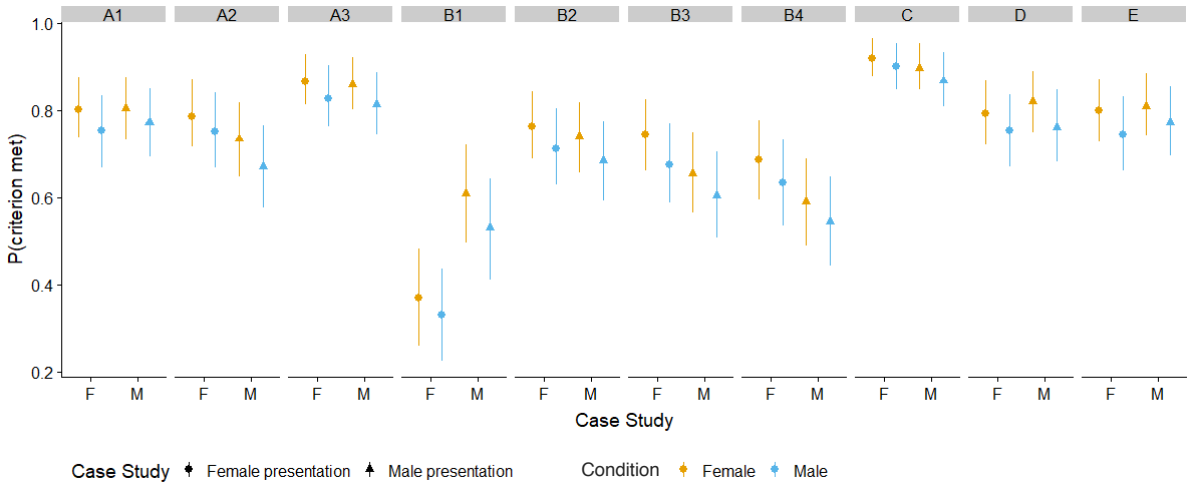
remaining dependent variables were numerical and, therefore, modelled as normally distributed and estimated as a linear combination of the predictors via an identity link function.

[59] Refer to Appendix L (Table L.1) for descriptive statistics on clinician endorsement of criteria, confidence and severity ratings across conditions.

confidence). The evidence did not allow for a precise conclusion at the 80% level as to whether there was an interaction between case study presentation and sex/gender condition (LOR = -0.01, $HDI_{80\%}$ = [-0.37, 0.34]; $P_{(\text{within ROPE})} = 36.1\% \approx P_{(\text{meaningful in either direction})}$).

Figure 5.2

Estimated Probability That Each Criterion Was Considered Met by Sex/Gender Condition and Case Study Presentation



Note. Criterion A1: deficits in social-emotional reciprocity; A2: deficits in nonverbal communication behaviours; A3: deficits in developing, maintaining, and understanding relationships; B1: stereotyped/ repetitive motor movements, use of objects, or speech; B2: insistence on sameness, routines, or ritualised behaviour; B3: restricted and fixated interests; B4: hyper-/hypo-reactivity to sensory input, C: symptoms present in the early developmental period; D: symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning; E: disturbances are not better explained by another diagnosis.

There was weak to moderate evidence that, across the two case studies, diagnosticians had a greater probability of considering any specific criterion met for the female sex/gender conditions (refer to Table 5.2). The strongest evidence of a difference in probability that a criterion was met for the male versus female sex/gender condition was found in Criterion A3

and the weakest in Criterion B4. However, in all cases, neither a negligible difference (nor a difference of zero) can be excluded with 80% confidence.

Table 5.2

Difference Between Sex/Gender Conditions in the Probability that Criteria Were Considered Met by Criterion and Collapsed Across Case Studies

Criterion	Log Odds Ratio [$HDI_{80\%}$]	$P_{(meaningful)}$
A1	-0.24 [-0.55, 0.03]	-.75
A2	-0.25 [-0.53, 0.04]	-.76
A3	-0.30 [-0.60, 0.01]	-.83
B1	-0.25 [-0.53, 0.03]	-.76
B2	-0.26 [-0.54, 0.03]	-.77
B3	-0.28 [-0.57, 0.00]	-.79
B4	-0.22 [-0.50, 0.07]	-.70
C	-0.26 [-0.57, 0.04]	-.75
D	-0.29 [-0.59, 0.00]	-.80
E	-0.26 [-0.56, 0.03]	-.77

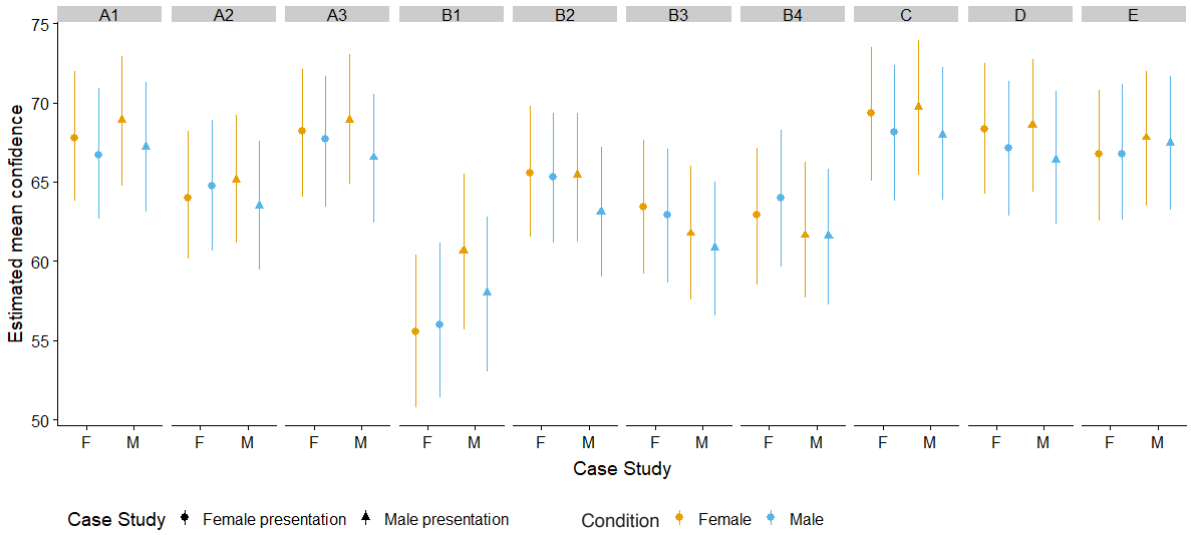
Note. $P_{(meaningful)}$ indicates the probability that the true difference fell outside the ROPE and in the observed direction (positive probabilities indicate greater probability for the male sex/gender conditions; negative probabilities for female sex/gender conditions).

Confidence that Criteria Were Met. The estimated overall effect of sex/gender condition on diagnosticians' confidence that the criteria were met was within the negligible range (d = -0.04, $HDI_{80\%}$ = [-0.12, 0.05], $P_{(within\ ROPE)}$ = 79.8%; see Figure 5.3). We can conclude with 80% confidence that diagnosticians' confidence that a criterion was met did not meaningfully vary between the male and female presentation case studies (d = -0.00, $HDI_{80\%}$ = [-0.09, 0.09]). Similarly, the balance of probabilities suggested that the interaction between case study presentation and sex/gender condition on confidence that criteria were met was negligible (d = -0.06, $HDI_{80\%}$ = [-0.21, 0.11], $P_{(within\ ROPE)}$ = 62.5%). Therefore, there

was no conclusive evidence that the effect of sex/gender condition on confidence that criteria were met differed by case study.

Figure 5.3

Estimated Confidence That Criteria Were Met by Sex/Gender Condition and Case Study



Note. Criterion A1: deficits in social-emotional reciprocity; A2: deficits in nonverbal communication behaviours; A3: deficits in developing, maintaining, and understanding relationships; B1: stereotyped/ repetitive motor movements, use of objects, or speech; B2: insistence on sameness, routines, or ritualised behaviour; B3: restricted and fixated interests; B4: hyper-/hypo-reactivity to sensory input, C: symptoms present in the early developmental period; D: symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning; E: disturbances are not better explained by another diagnosis.

The evidence was equivocal as to whether diagnosticians' confidence that any particular criterion was met differed meaningfully according to sex/gender condition (across the case studies; refer to Table 5.3). In all cases, the evidence did not clearly support either an effect or no difference, but the balance of probabilities was in favour of no effect.

Table 5.3

Difference Between Sex/Gender Conditions in the Confidence That Criteria Were Met by Criterion and Collapsed Across Case Studies

Criterion	*d* [$HDI_{80\%}$]	$P_{(meaningful)}$	$P_{(within\ ROPE)}$
A1	-0.05 [-0.18, 0.07]	-.32	.63
A2	-0.02 [-0.15, 0.11]	-.20	.63
A3	-0.05 [-0.18, 0.08]	-.32	.63
B1	-0.04 [-0.17, 0.09]	-.27	.65
B2	-0.05 [-0.18, 0.08]	-.30	.64
B3	-0.03 [-0.16, 0.09]	-.23	.67
B4	0.01 [-0.14, 0.15]	.22	.65
C	-0.05 [-0.19, 0.07]	-.33	.63
D	-0.06 [-0.20, 0.07]	-.35	.61
E	-0.01 [-0.15, 0.12]	-.18	.67

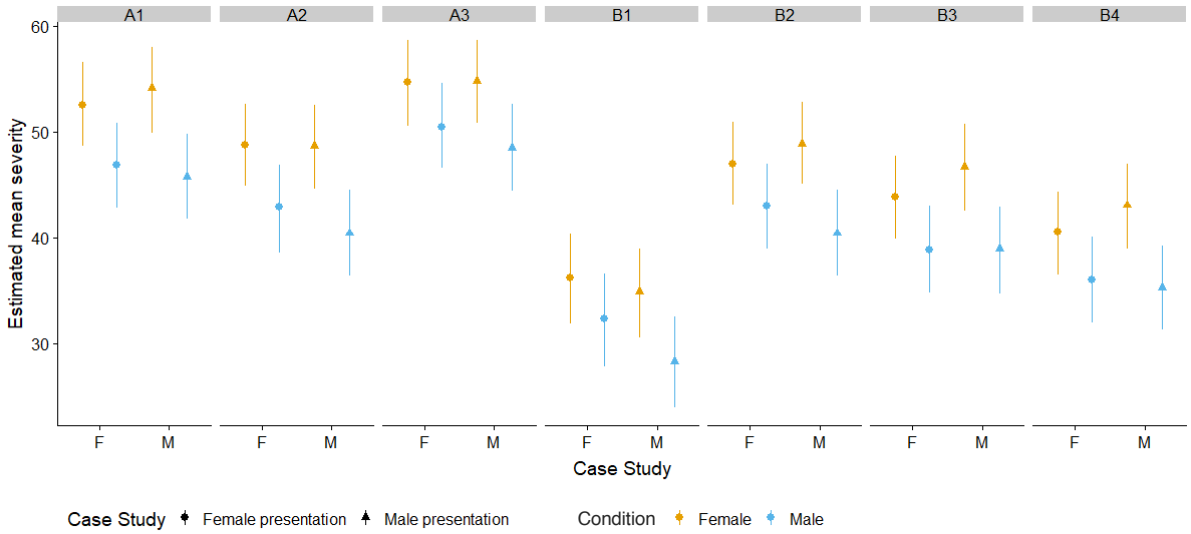
Note. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction (positive probabilities indicate greater probability for the male sex/gender conditions; negative probabilities for female sex/gender conditions). $P_{(within\ ROPE)}$ = probability that the true difference lay within the ROPE.

Perceived Severity of Difficulties. Despite no evidence of a difference in confidence that criteria were met and equivocal evidence for the probability that criteria were considered met across sex/gender conditions, hierarchical Bayesian regression showed a meaningful association between sex/gender conditions and diagnosticians' rating of the severity of difficulties across Criteria A and B (refer to Figure 5.4). Within each case study, the severity of difficulties was consistently rated as greater in the female sex/gender condition (d = -0.28, $HDI_{80\%}$ = [-0.41, -0.16]; $P_{(meaningful)}$ = 96.2%). Evidence suggested no meaningful difference between case studies in whether difficulties were rated as more severe in the female sex/gender conditions. However, a small difference cannot be excluded with 80% confidence (d = -0.02, $HDI_{80\%}$ = [-0.12, 0.09], $P_{(within\ ROPE)}$ = 90.5%). Finally, the interaction between case

study and sex/gender condition on severity ratings was equivocal (d = -0.13, $HDI_{80\%}$ = [-0.34, 0.07], $P_{(within\ ROPE)}$ = 41.5%). Therefore, there was no conclusive evidence that the effect of sex/gender condition on severity ratings differed by case study.

Figure 5.4

Estimated Perceived Severity of Difficulties by Sex/Gender Condition and Case Study



Note. Criterion A1: deficits in social-emotional reciprocity; A2: deficits in nonverbal communication behaviours; A3: deficits in developing, maintaining, and understanding relationships; B1: stereotyped/ repetitive motor movements, use of objects, or speech; B2: insistence on sameness, routines, or ritualised behaviour; B3: restricted and fixated interests; B4: hyper-/hypo-reactivity to sensory input.

As Figure 5.4 indicates, diagnosticians reported higher severity ratings in the female sex/gender condition across all criteria in both case studies. However, these differences were least apparent in Criteria A3 and B1. Table 5.4 presents effect sizes for the differences between sex/gender conditions in perceived severity for each criterion, averaged across the case studies. For each criterion, the effect size was small and generally similar across criteria.

Table 5.4

Difference Between Sex/Gender Conditions in the Perceived Severity of Difficulties by Criterion and Collapsed Across Case Studies

Criterion	*d* [$HDI_{80\%}$]	$P_{(meaningful)}$
A1	**-0.32 [-0.48, -0.17]**	**-.97**
A2	**-0.32 [-0.48, -0.17]**	**-.97**
A3	-0.25 [-0.40, -0.09]	-.87
B1	-0.25 [-0.41, -0.09]	-.87
B2	**-0.29 [-0.44, -0.14]**	**-.94**
B3	**-0.29 [-0.44, -0.14]**	**-.95**
B4	**-0.29 [-0.43, -0.14]**	**-.94**

Note. Rows in boldface indicate that the $HDI_{80\%}$ fell entirely outside the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction (positive probabilities indicate greater probability for the male sex/gender conditions; negative probabilities for female sex/gender conditions).

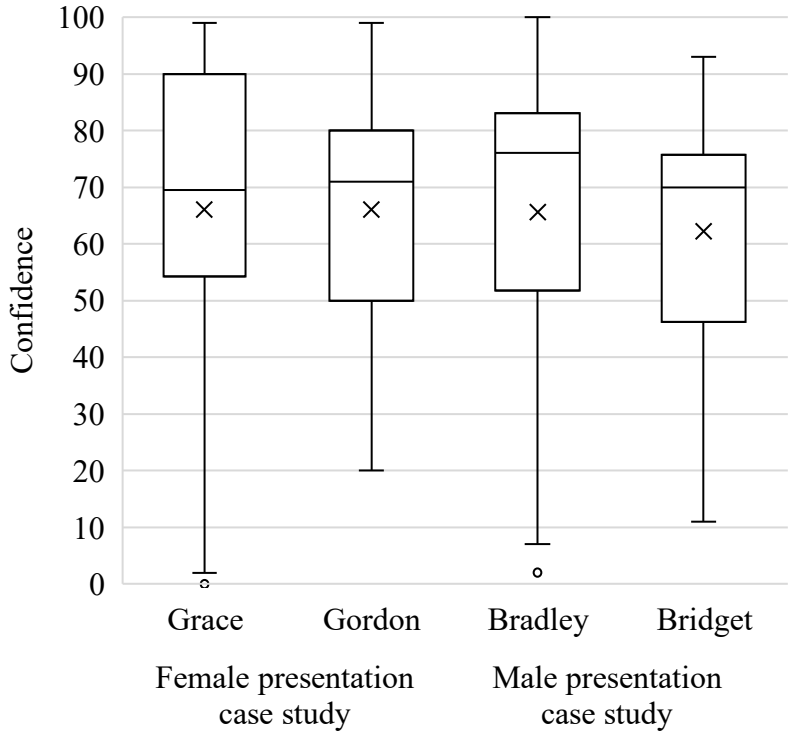
Confidence in ASD Diagnosis

Diagnosticians' confidence that the child described in each case study presented with ASD was measured on a scale from 0 (*not at all confident*) to 100 (*extremely confident*; refer to Figure 5.5). For the female presentation case study, Bayesian *t*-tests revealed no evidence of an effect of sex/gender condition on diagnosticians' confidence that the child had ASD (Grace; $M = 67.65$, $HDI_{80\%} = [60.27, 75.41]$, $SD = 26.40$ and Gordon; $M = 67.07$, $HDI_{80\%} = [60.27. 73.90]$, $SD = 21.97$). The evidence did not allow for a precise conclusion at the 80% confidence level ($d = -0.03$, $HDI_{80\%} = [-0.42, 0.42]$, $P_{(within\ ROPE)} = 23.8\%$; $P_{(meaningful)} = 38.1\%$, both above and below ROPE). Similarly, there was equivocal evidence of a difference in diagnosticians' confidence in an ASD diagnosis between conditions in the male presentation case study (Bradley; $M = 67.16$, $HDI_{80\%} = [59.24, 76.24]$ $SD = 27.35$ and Bridget; $M = 63.02$, $HDI_{80\%} = [56.27, 70.16]$, $SD = 22.81$; $d = 0.17$, $HDI_{80\%} = [-0.27, 0.59]$; $P_{(meaningful)} = 57.6\%$,

$P_{(within\ ROPE)}$ = 21.3%). There was a large degree of uncertainty in each of these comparisons as a result of the considerable, and perhaps surprising, variability in diagnosticians' confidence in an ASD diagnosis within each condition.

Figure 5.5

Box and Whisker Plot of Diagnosticians' Confidence That the Child Presents with ASD



Note. Crosses indicate means.

Levels of Support Required

Diagnosticians assigned support levels for social communication (Criteria A) and repetitive and restricted behaviours (Criteria B) for each case study. Support levels were defined according to the DSM-5 (American Psychiatric Association, 2013): *Level 1: Requiring Support*; *Level 2: Requiring Substantial Support*; and *Level 3: Requiring Very Substantial Support*. Due to the small number of Level 3 selections, Levels 2 and 3 were combined into a single category for the logistic regression analysis.

As shown in Table 5.5, there was no evidence of meaningful differences in the probability that the male or female sex/gender condition would be assigned any support level in either case study.[60] The effect sizes were negligible, but estimated with sufficient uncertainty to prevent ruling out a difference in either direction.

Table 5.5

Results of Logistic Regression for Allocated Support Levels

Support Level	Female presentation case study		Male presentation case study	
	Log Odds Ratio [$HDI_{80\%}$]	$P_{(meaningful)}$	Log Odds Ratio [$HDI_{80\%}$]	$P_{(meaningful)}$
None	-0.01 [-0.17, 0.12]	-.44	0.04 [-0.10, 0.21]	.54
Level 1	0.06 [-0.11, 0.26]	.61	-0.06 [-0.25, 0.11]	-.62
Level 2/3	-0.04 [-0.24, 0.13]	-.56	0.02 [-0.16, 0.19]	.47

Note. Level 1 = Requiring Support, Level 2 = Requiring Substantial Support, Level 3 = Requiring Very Substantial Support, None = insufficient criteria met. Positive log odds ratios indicate that males had a greater probability of being assigned a particular support level. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction (positive probabilities indicate greater probability for male sex/gender condition; negative probabilities for female sex/gender condition).

Diagnosis

Diagnosticians' impressions regarding diagnoses other than ASD were also considered and compared between conditions in each case study. Specifically, Bayesian logistic regressions were used to assess which diagnoses meaningfully predicted sex/gender

[60] Refer to Appendix L (Figure L) for descriptive data about the frequency of support level selections.

condition for each case study. An identical process was used for the analysis of additional or differential diagnoses that diagnosticians would consider exploring further.

As shown in Table 5.6, two diagnostic categories were found to meaningfully predict sex/gender condition: *Attention-deficit/hyperactivity disorder (ADHD)* and *No diagnosis*.[61] This evidence was robust ($P_{(meaningful)} \geq 98\%$) and there was a higher probability that each category would be assigned in the male sex/gender conditions (indicated by the 'Sex/Gender Condition' column of Table 5.6). Although results suggest moderate evidence that *Developmental delay* and *Intellectual disability* were more likely to be endorsed for the female sex/gender conditions ($P_{(meaningful)} = 80\text{-}81\%$), this must be interpreted with the proviso that these diagnoses were not frequently selected (i.e., three or less instances across all conditions for each diagnosis). As the interaction coefficients were not meaningful for any of the above diagnoses, there is no evidence of a difference in the size of these effects across the male and female presentation case studies.

Two additional or differential diagnoses (offered either in conjunction with or instead of ASD diagnosis) meaningfully predicted sex/gender condition across case studies: ADHD and generalised anxiety disorder (GAD), with a higher probability that the diagnostician would consider these diagnoses in the male sex/gender conditions. This evidence was also robust ($P_{(meaningful)} \geq 93\%$). There is no evidence of a difference in the size of this effect across the male and female presentation case studies.

Another five additional/differential diagnoses were presented as options for further consideration: *Separation anxiety disorder*, *Reactive attachment disorder*, *Post-traumatic stress disorder*, *Conduct disorder*, and *Oppositional defiant disorder*, and a further four were

[61] Refer to Appendix L for descriptive statistics on endorsement of diagnoses (Table L.2) and supplementary logistic regression statistics (i.e., effect of case study and interaction terms; Table L.3).

Table 5.6

Results of Logistic Regression for Diagnoses by Sex/Gender Condition

	Intercept	Sex/Gender Condition	
	Log Odds Ratio [$HDI_{80\%}$]	Log Odds Ratio [$HDI_{80\%}$]	$P_{(meaningful)}$
Diagnosis			
ASD	1.38 [1.01, 1.73]	-0.31 [-1.05, 0.39]	-.65
ADHD	-2.68 [-3.42, -1.80]	**2.70 [-.95, 4.11]**	**1.00**
Generalised anxiety disorder	-2.72 [-3.27, -1.80]	-0.44 [-1.60, 0.81]	-.64
Developmental delay	-4.34 [-5.34, -3.09]	-1.56 [-3.54, 0.76]	-.80
Intellectual disability	-7.32 [-9.41, -4.66]	-2.76 [-6.35, 1.41]	-.81
No diagnosis	-2.91 [-3.64, -1.95]	**2.35 [0.45, 3.75]**	**.98**
Additional/differential diagnoses			
ADHD	0.52 [0.20, 0.86]	**0.87 [0.19, 1.52]**	**.93**
Generalised anxiety disorder	-0.96 [-1.36, -0.49]	**1.32 [0.39, 2.12]**	**.97**
Social anxiety disorder	-1.38 [-1.80, -0.95]	-0.59 [-1.45, -0.22]	-.78
Obsessive-compulsive disorder	-2.66 [-3.21, -2.04]	-0.53 [-1.68, 0.64]	-.67
Social communication disorder	-0.37 [-0.68, -0.08]	0.22 [-0.37, 0.83]	.59
Language disorder	-1.19 [-1.54, -0.84]	0.20 [-0.50, 0.90]	.57
Intellectual disability	-1.35 [-1.69, -0.96]	-0.39 [-1.11, 0.35]	-.69

Note. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. Underlined coefficients highlight variables predictive of female sex/gender condition. Positive log odds ratios indicate that males had a greater probability of being assigned a particular diagnosis. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction (positive probabilities indicate greater probability of disorder endorsement for male sex/gender condition; negative probabilities for female sex/gender condition).

not supplied as options but entered by diagnosticians under *Other (specified): Specific learning disorder*, *Sensory processing disorder*, *Developmental delay*, and *Foetal alcohol syndrome*. These diagnoses were included in the logistic regression analysis but, as these options were not often selected (between one and six times across case studies), the HDIs

$_{(80\%)}$ were too wide to interpret and no valid conclusions could be drawn about the strength with which their selection was predicted by the sex/gender condition.

Subjective Difficulty Ratings

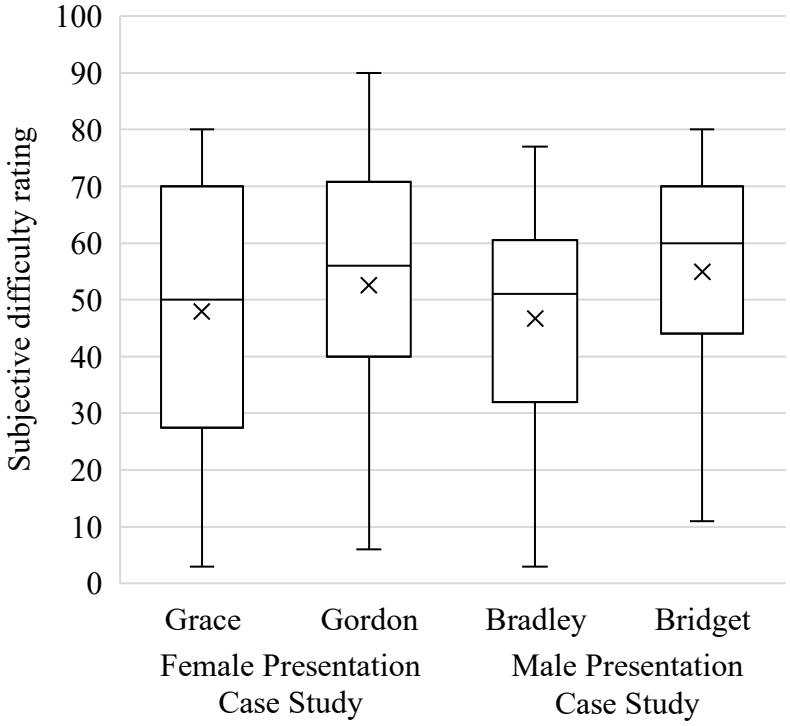
Diagnosticians rated the difficulty of reaching a diagnostic conclusion for each case study from 0 (*extremely easy*) to 100 (*extremely difficult*) and data were analysed using Bayesian *t*-tests (see Figure 5.6). Broadly, diagnosticians' subjective difficulty ratings for reaching a diagnosis in both case studies tended towards greater difficulty when the sex/gender condition was incongruent with the presentation (e.g., male sex/gender condition for the female presentation case study). There was no evidence of a meaningful difference in subjective difficulty ratings between sex/gender conditions in the female presentation case study (Grace; $M = 48.50$, $HDI_{80\%} = [40.86, 55.97]$, $SD = 26.89$ and Gordon; $M = 52.98$, $HDI_{80\%} = [45.69, 60.31]$, $SD = 25.03$; M difference = 4.49). Neither a small difference in either direction, nor a negligible difference, can be excluded based on these data ($d = 0.17$, $HDI_{80\%} = [-0.23, 0.58]$). We have only 59.2% confidence that diagnosticians had meaningfully greater difficulty reaching a diagnosis for Gordon ($P_{(\text{within ROPE})} = 21.2\%$).

In contrast, there was evidence of a probable difference in subjective difficulty ratings between conditions in the male presentation case study (Bradley; $M = 47.39$, $HDI_{80\%} = [41.73, 53.35]$, $SD = 19.84$ and Bridget; $M = 55.74$, $HDI_{80\%} = [49.85, 61.68]$, $SD = 19.04$; M difference = 8.35). We can have 83.7% confidence that diagnosticians had meaningfully greater difficulty reaching a diagnostic conclusion for Bridget ($d = -0.44$, $HDI_{80\%} = [-0.89, -0.02]$).[62]

[62] A Bayesian correlation coefficient test revealed a moderate positive correlation between diagnosticians' ratings of subjective difficulty for each of the male and female presentation case studies ($\rho = 0.43$, $HDI_{80\%} = [0.26, 0.61]$). We can be 98.0% confident that the true correlation is positive and meaningful.

Figure 5.6

Box and Whisker Plot of Subjective Difficulty to Reach Diagnosis for Each Presentation and Condition



Note. Crosses indicate means. 0 = *Extremely easy*; 100 = *Extremely difficult*.

Experience Assessing Females for ASD

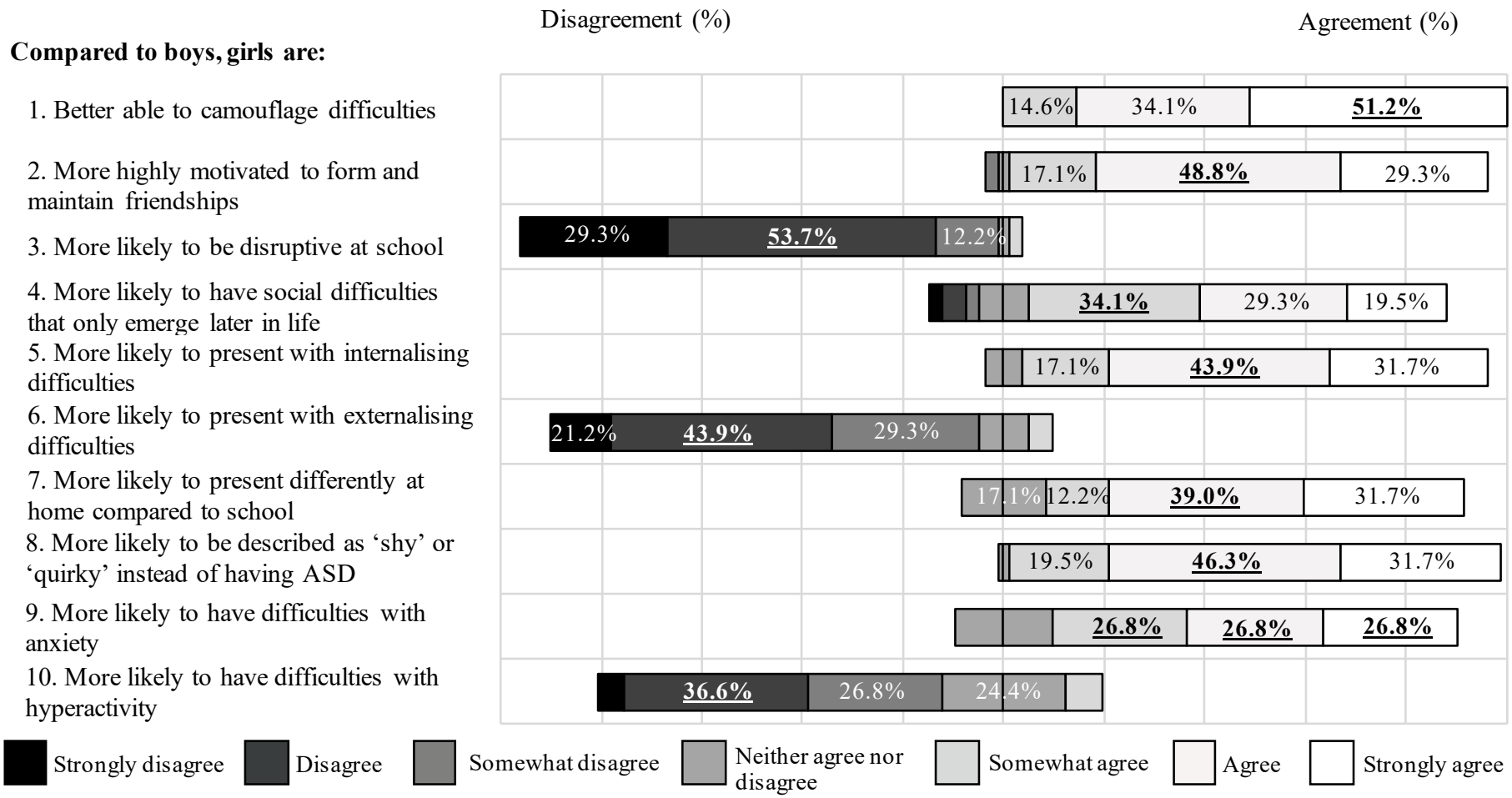
Do Diagnosticians Perceive Sex/Gender Differences in the Presentation of ASD?

All diagnosticians reported that there are differences in how males and females with ASD typically present. Of these, 7.3% (n = 3) rated these differences as *subtle*, whereas 43.9% (n = 18) rated them as *moderate* and 48.8% (n = 20) rated them as *marked*.

Diagnosticians were presented with 10 statements about possible sex/gender differences in ASD and seven-point Likert scales on which they rated the extent of their agreement from 1 (*strongly disagree*) to 7 (*strongly agree*). Figure 5.7 presents the frequency of all selections. There was generally strong agreement with the statements that, compared

Figure 5.7

Plot of Likert Scale Endorsed Items (% of Diagnosticians, n = 41) by Item (1-10)



with boys, girls with ASD are better able to camouflage difficulties (Item 1), and agreement that girls are more highly motivated to form and maintain friendships (Item 2), more likely to present with internalising difficulties (Item 5) and present differently at home compared to school (Item 7). Diagnosticians also generally agreed with the statements that girls are more likely than boys to be described as 'shy' or 'quirky' instead of having ASD (Item 8) and have difficulties with anxiety (Item 9). There was weak agreement with the statement that girls are more likely than boys to have social difficulties that only emerge later in life (Item 4). Conversely, there was general disagreement with reverse coded items: i.e., that compared to boys, girls are more likely to be disruptive at school (Item 3), present with externalising difficulties (Item 6) and have difficulties with hyperactivity (Item 10).

Content analysis was used to interpret diagnosticians' responses to the question, "What are three features of ASD that may present differently in females?" Categories were coded according to the DSM-5 ASD criteria (see Table 5.7). As highlighted in the table, surveyed diagnosticians reported both quantitative and qualitative differences in presentation across the ASD criteria, but primarily in the social communication domain and Criterion B3. Differences were not commonly reported in criteria B1, B2, or B4, but were in strategies used to manage ASD symptoms (e.g., masking, internalising/ externalising behaviour).

Table 5.7

Features of ASD That Present Differently in Females (n = 35 Diagnosticians, 101 responses)

Category	Responses *(n)*	Example Quote
Criteria A (DSM-5)	**Total: 56**	
Social communication	14	"[Girls have] better social skills than boys."
A1. Social-emotional reciprocity	13	"Increased likelihood of social approach - but it will be awkward. Female autistics [sic] seem to me to be more likely to be indiscriminately social and at high risk of *stranger danger*."
A2. Non-verbal communication	16	"Girls present with better non-verbals generally e.g., eye contact, gestures and facial expression."
A3. Developing, maintaining & understanding relationships	13	"Girls often seem to have a group of friends but will be on the outer rather than being completely alone, making it look like they have friends."
Criteria B (DSM-5)	**Total: 30**	
Restricted and repetitive behaviours and interests	3	"[Girls present with] more subtle repetitive behaviour."
B1. Stereotyped/repetitive behaviour	2	"Females may have fewer stereotypies."
B2. Insistence on sameness, rituals and routines	4	"Rigidity [is] often more subtle and in viewpoint and opinions rather than in something external that can be more easily observed."
B3. Restricted interests	20	"[Girls'] interests are less unusual and challenges in repetitive and restricted interests may be more subtle and go unnoticed (e.g., will order dolls rather than engage in pretend play)."
B4. Sensory sensitivity	1	"[Girls have] different sensory difficulties."
Other	**Total: 15**	
Masking difficulties	11	"Social difficulties may be masked by superficial coping techniques e.g., mimicry of others at school."
Internalising/externalising behaviour	3	"Females perhaps internalise responses to difficulties more than they externalise."
Energy level	1	"Females [are] more hypoactive as opposed to hyperactive."

Note. Not all diagnosticians responded to this question and some provided fewer than three responses.

Clinician Familiarity and Confidence

There was substantial variability in diagnosticians' self-rated familiarity with differences in how ASD may present between males and females. On a scale of familiarity ranging from 0 (*not at all familiar*) to 100 (*very familiar*), diagnosticians rated themselves at 67.51 on average (*SD* = 18.59). Approximately one-third rated their familiarity above or equal to 80 (*n* = 15, 36.6%) and a minority (*n* = 10, 24.4%) rated their familiarity below or equal to 50. Diagnostician familiarity with sex/gender differences in ASD presentation was weakly positively correlated with the length of experience working with people with ASD (ρ = 0.23, $HDI_{80\%}$ = [0.05, 0.45], $P_{(meaningful)}$ = 79.2%) and, as one might expect, more strongly with length of experience conducting assessments for ASD (ρ = 0.35, $HDI_{80\%}$ = [0.18, 0.55], $P_{(meaningful)}$ = 95.3%).[63]

A Bayesian ANOVA was used to compare diagnosticians' self-reported confidence in accurately diagnosing different client groups presenting for assessment query ASD. Table 5.8 contains the results of this analysis and Figure 5.8 presents the comparisons graphically. Only diagnosticians who reported assessing at least one person from each client group for ASD during a six-month period were included in these comparisons.

Compared to girls (aged under 18 years), men, and women, diagnosticians had greatest confidence in accurately diagnosing boys (aged under 18 years). There was robust evidence suggesting that diagnosticians were more confident in assessing boys compared to all other groups, but particularly compared to women and girls. There was weak evidence that diagnosticians had greater confidence in assessing girls than women, although neither a negligible difference nor a difference of zero could be excluded with 80% confidence.

[63] Familiarity with sex/gender differences in ASD presentation was moderately negatively correlated with subjective difficulty in reaching a diagnosis for the female presentation case study (ρ = -0.51, $HDI_{80\%}$ = [-0.68, -0.36], $P_{(meaningful)}$ = 99.6%).

Finally, there was limited evidence of a difference in diagnostician confidence assessing men compared to girls.

Table 5.8

Diagnosticians' Confidence in Accurately Diagnosing Each Client Group Presenting Query ASD (Adults; n = 25, Children; n = 41 Responses)

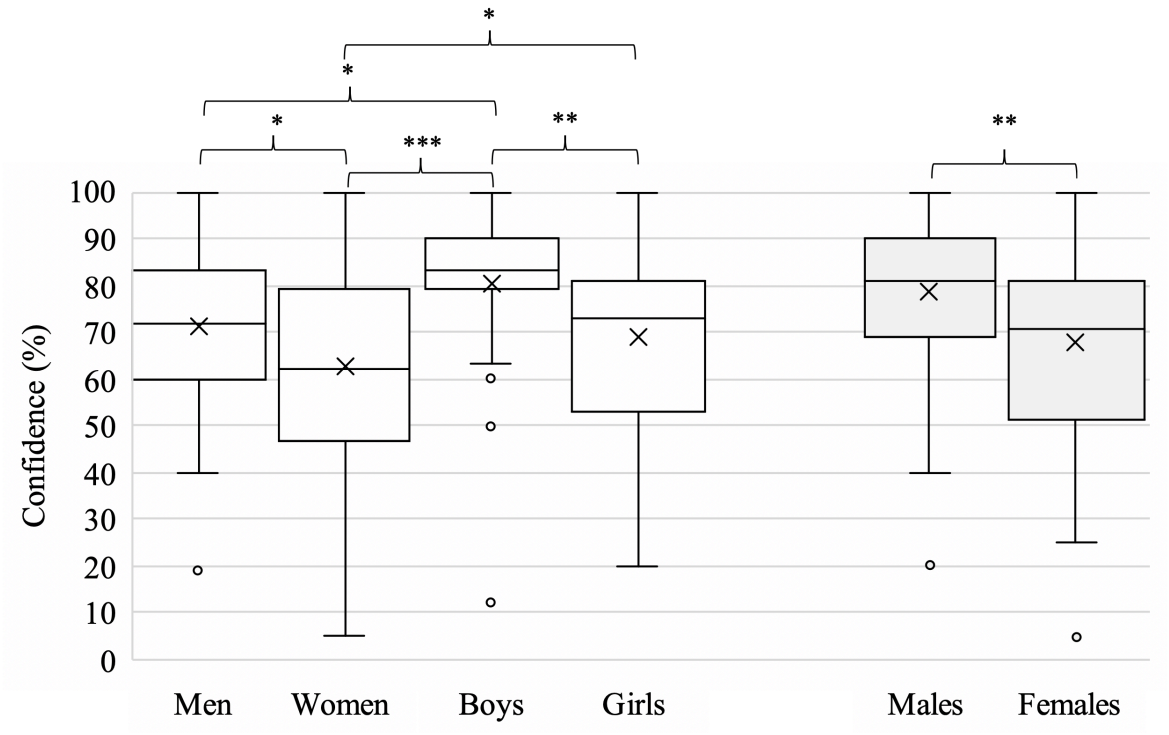
Client group			Women	Boys	Girls
	Confidence *M* [$HDI_{80\%}$] *(SD)*		65.58 [61.34, 69.89] (22.49)	81.49 [77.99, 85.05] (11.56)	70.72 [67.32, 74.40] (19.71)
Men			Men > women	Boys > men	Men > girls
	73.66 [69.46, 77.89] (18.53)		**0.43** **[0.12, 0.73]** $P_{(meaningful)}$ = .90	**0.41** **[0.11, 0.69]** $P_{(meaningful)}$ = .92	0.16 [-0.14 0.48] $P_{(meaningful)}$ = .59
Women		Direction *d* [$HDI_{80\%}$] $P_{(meaningful)}$		Boys > women	Girls > women
	65.58 [61.34, 69.89] (22.49)			**0.84** **[0.52, 1.15]** $P_{(meaningful)}$ = 1.00	0.27 [-0.02, 0.56] $P_{(meaningful)}$ = .78
Boys					Boys > girls
	81.49 [77.99, 85.05] (11.56)				**0.56** **[0.31, 0.82]** $P_{(meaningful)}$ = .99

Note. Confidence was rated on a scale from 0 (*not at all confiden*t) to 100 (*extremely confident*). Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction.

A Bayesian *t*-test was conducted on the aggregation of data for males (boys and men; *M* = 79.18, $HDI_{80\%}$ = [76.66, 81.58], *SD* = 14.88) and females (girls and women; *M* = 68.20, $HDI_{80\%}$ = [64.87, 71.63], *SD* = 20.51). Consistent with the results presented above, there was robust evidence ($P_{(meaningful)}$ = 99.7%) that diagnosticians were meaningfully more confident in accurately diagnosing males than females (*d* = -0.62, $HDI_{80\%}$ = [-0.85, -0.37]).

Figure 5.8

Diagnosticians' Mean Confidence in Accurately Diagnosing Each Client Group Presenting Query ASD (Adults; n = 25, Children; n = 41 Responses)



Note. * small-medium effect (*d* between 0.2 and 0.5), ** medium effect (0.5 to 0.8), *** strong effect (> 0.8; Cohen, 1988). Crosses indicate means. 0 = *Not at all confident;* 100 = *Extremely confident*.

Challenges in Assessing Females for ASD

Overall, the majority of diagnosticians reported that they perceive ASD assessment to be more challenging when the client is female (87.8%), while the remaining minority reported that assessment is equally challenging regardless of the client's sex/gender. Of diagnosticians who deemed it more challenging to assess a female client for ASD (*n* = 36), 22.2% reported that male clients are *much easier*, 44.4% that males are *moderately easier,* and 33.3% that males are *slightly easier* to assess.

Diagnosticians' responses to the questions, "what are three reasons that females with ASD may be underdiagnosed?" and, "what are the most challenging aspects of assessing a

female for ASD?" were analysed together. The code structure for responses is presented in Table 5.9. [64] Most notably, the current ASD instruments and conceptualisation were appraised as being incongruent with the presentation of females, and therefore diagnosticians reported that they rely upon their own clinical judgement to determine whether a female reaches the threshold for ASD diagnosis.

Table 5.9

Reasons for Underdiagnosis and Greatest Challenges in Assessing Females for ASD (n = 40 Diagnosticians, 177 Responses)

Category	Responses *(n)*	Example Quote
1. Negotiating the mismatch in symptom manifestation with 'classic' ASD	122	
(a) Camouflaging: Masking and imitating	55	"Masking is often so well developed that [females] can be very *neurotypical-passing*. Asking the right questions to understand how hard they have to work to pass as neurotypical can be challenging."
(b) Bias in ASD conceptualisation and assessment instruments	24	"I lack trust in scores on instruments and tools when working with females, knowing they've been developed with a male bias… I'm not sure there exists a clear conceptualisation of autism in females, meaning some girls might be missed."
(c) More subtle presentation	15	"Characteristics are more subtle or have less of a functional impact."
(d) More socially and developmentally appropriate restricted interests	12	"Their interests may be more socially acceptable… whereas boys' [interests] are less mainstream."

[64] The category structures are presented separately in Appendix M (Table M.1 and M.2).

Category	Responses *(n)*	Example Quote
(e) Less disruptive/ externalising behaviour	9	"In my experience, girls typically present as *internalisers* rather than frequently externalising children who attract adult attention."
(f) Greater social motivation	7	"Increased interest in peers/desire for friendships (just lacking in skills to develop and maintain them)."
2. The female presentation of ASD remains poorly understood	18	
(a) Amongst clinicians and diagnosticians	9	"Different professionals have different views on whether sex differences exist."
(b) In the community	5	"Lack of awareness in community: [girls are] less likely to be referred."
(c) In research	4	"Limited empirical research into presentation of ASD in females."
3. Misdiagnosis or diagnostic overshadowing, especially with anxiety disorders	15	"Often female autistics [sic] seem to be misdiagnosed with social anxiety, borderline personality disorder or another psychiatric illness. Teasing apart these conditions can be challenging, especially since almost all female autistics have comorbid anxiety and/or depression as a result of their social difficulties."
4. Females' presentations may vary across different environments and with time	11	"Data across environments is often at odds… especially when girls are able to 'hold onto' their reactions at school. They typically present when social environment exceeds capacity, which I find is often 8+ years, thus diagnosed later than [the] average male."
5. Professionals do not look for ASD in females: Sex/gender expectations and priming	7	"Professionals have pre-conceptions of [the] autism 'presentation'."
6. Normative sex/gender differences exist in neurology and developmental trajectories	4	"[Girls] may have a developmental advantage in social engagement."

Note. Not all diagnosticians responded to these questions and some provided fewer than six responses.

Circumventing Challenges

Codes for diagnosticians' responses to the question, "What have you changed about your assessment procedures to circumvent challenges assessing females for ASD?" are presented in Table 5.10.

Table 5.10

Categories Identified for Circumventing Challenges in Assessing Females for ASD (n = 24 Diagnosticians, 44 Responses)

Category	Responses *(n)*	Example Quote
1. Investing time to develop an appropriate case conceptualisation	8	"Increased time/sessions devoted to the assessment process to ensure increased personal understanding of presentation of the client."
2. Adapting the parent interview: Using specific and targeted questioning	7	"More prompting and probing during interviews with parents and asking very specific questions."
3. Negotiating multiple sources of information	7	"More reliance on parent report and my own observations."
4. Recognising limitations in assessment instruments	7	"Understanding [that] diagnostic assessment tools, e.g., ADOS, may not be sensitive to all signs and symptoms."
5. Conducting more comprehensive behaviour observation across multiple settings	5	"Using imaginative play and specific assessment tools to gain more observations (such as the use of picture stimuli to extract responses to social situations)."
6. Collecting self-report information	5	"I ask questions to try to ascertain whether they are actively 'masking.'"
7. Improving understanding of female presentation amongst professionals/ parents	4	"We have all learnt about the differences in presentation and so actively view girls through a different lens."
8. Considering features outside ASD criteria	1	"Looking more closely at the constellation of characteristics outside of the criteria that might be present (e.g., gut and sleep issues)."

Note. Not all diagnosticians responded to this question.

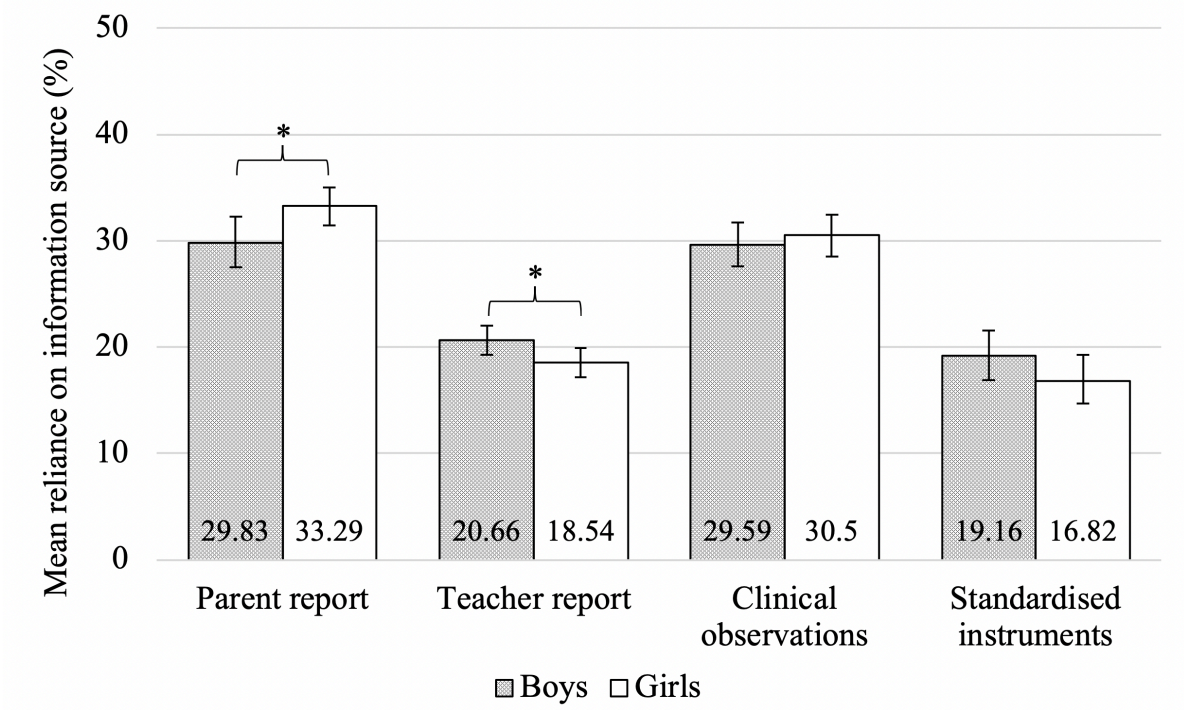
Finally, Bayesian *t*-tests were used to examine the relative strength with which diagnosticians weighted different sources of diagnostic information between boys and girls (i.e., parent report, teacher report, clinical observations, and the results of standardised psychometric instruments) in order to further clarify their existing methods for circumventing challenges in the ASD assessment of females.

Averaged across their ASD assessments for males and females, diagnosticians reported relying most heavily on parent report (M = 31.57%, SD = 10.41%) and clinical observations (M = 30.05%, SD = 9.70%) in forming a diagnostic opinion. They reported relying less heavily upon teacher report (M = 19.60%, SD = 6.66%) and the results of standardised instruments (M = 17.98%, SD = 11.34%).

No strong evidence was found for a meaningful difference in diagnosticians' reliance on any given source of information by client sex/gender (see Figure 5.9). However, there was moderate evidence ($P_{(meaningful)}$ = 86.0%) that, for females, diagnosticians place meaningfully greater emphasis on parent report (d = -0.35, $HDI_{80\%}$ = [-0.66, -0.06]) and less emphasis on teacher report (d = 0.32, $HDI_{80\%}$ = [0.02, 0.62], $P_{(meaningful)}$ = 83.1%) compared to males. Equivocal evidence was found for differences in reliance upon clinical observations (d = -0.10, $HDI_{80\%}$ = [-0.38, 0.20]) and standardised instruments (d = 0.21, $HDI_{80\%}$ = -0.10, 0.50]) between males and females. For clinical observations, the balance of probabilities suggested a 30.8% probability of a negligibly sized difference and an effect in either direction could not be excluded with 80% confidence. For the results of standardised instruments, analysis indicated a 68.0% probability that diagnosticians rely more strongly on this source for males than females ($P_{(within\ ROPE)}$ = 22.8%).

Figure 5.9

Mean Reliance on Information Sources in ASD Assessment by Sex/Gender (n = 41)



Note. Error bars indicate HDIs $_{(80\%)}$. * moderate evidence of meaningful difference (i.e., a difference of zero, but not a negligibly sized difference, can be excluded with 80% confidence).

Discussion

This is among the first studies to examine sex/gender expectancy bias in diagnostic decision-making for ASD. Specifically, I aimed to identify why diagnostic assessment may be more challenging for female clients and how future research might be oriented to address these challenges. ASD diagnosticians were presented with two case studies (reflecting a 'classic' male ASD presentation and a female presentation, respectively) with the sex/gender of the child randomly allocated within each. Sex/gender condition was not found to be meaningfully associated with diagnosticians' confidence that criteria were met or the probability of ASD diagnosis in either case study (irrespective of congruence with the sex/gender of the case study presentation). However, ASD-related behaviours were consistently rated as more severe in the female sex/gender conditions, across both case

studies. Diagnosticians identified a large number of challenges in assessing females for ASD, many of which related to phenotypic sex/gender differences and identifying camouflaging behaviours. Furthermore, they suggested strategies for overcoming such challenges and identified priorities for future research.

Sex/Gender Expectancy Bias

Sex/gender expectancy bias has previously been demonstrated among diagnosticians in a number of psychiatric disorders (Hartung & Widiger, 1998). As ASD is more prevalent and diagnosed more often in males and its features are considered more typical of males (the core tenet of the Extreme Male Brain theory; Baron-Cohen & Hammer, 1997), it was expected that diagnosticians would be more likely to endorse an ASD diagnosis when male pseudonyms were assigned in the case studies. However, there was no meaningful effect of sex/gender condition on (a) the likelihood of ASD diagnosis, nor (b) confidence that any criterion was met, in either case study presentation. These findings contrast with previous literature suggesting that females may be less likely than males to receive an ASD diagnosis at assessment, even if they present with equally severe ASD traits (Russell et al., 2011; Wilson et al., 2016).

There are three main reasons that these results may contrast with those of previous investigations. First, as a result of the increased empirical and media attention this topic has received over recent years, diagnosticians recruited in this study may have had greater awareness of the female ASD presentation, heterogeneity of presentations, and different ways that ASD criteria may be met, compared to diagnosticians who contributed to previous studies. A second possibility, and an important limitation of this study, is that the recruitment materials specified females with ASD as its focus. The sample may therefore be more reflective of diagnosticians with greater confidence or knowledge in assessing females for ASD, or who are otherwise primed to be ‘on the lookout’ for ASD characteristics among

females. Finally, demand characteristics associated with the advertised focus of the study may have influenced diagnosticians to respond with greater sensitivity to the ASD-related difficulties of females. It is unknown what effect, if any, priming and demand characteristics had upon the clinical decision-making examined.

While sex/gender condition had no meaningful effect on diagnosticians' confidence in an ASD diagnosis or specific ASD criteria being met, diagnosticians perceived ASD symptoms as being more severe in the female sex/gender conditions. In other words, the same ASD symptoms were perceived as more severe for the female sex/gender conditions, but this did not render diagnosticians more certain of an ASD diagnosis.[65] The current diagnostic instruments and disorder conceptualisation are typically based upon a single threshold at which ASD traits are regarded *clinically significant* and therefore qualifying for formal diagnosis. However, the results of this study suggest that diagnosticians may have different perceptions of severity thresholds for males and females, at which they are equally certain of an ASD diagnosis, and against which they determine whether a criterion is met. Studies of sex/gender differences in typically developing children have suggested that, on average, boys display higher levels of ASD traits than girls (Constantino & Todd, 2003) and girls may demonstrate an early advantage in social skill development (Kreiser & White, 2014; Leman & Tenenbaum, 2011). Therefore, greater ASD difficulty may be required of girls in order to meet the threshold for ASD. Conversely, when boys' and girls' presentations are identical (as in the present study), girls' presentations may be considered more atypical and hence more severely autistic than boys. This mirrors previous findings (e.g., Dworzynski

[65] There was some evidence that overall, and in some criteria in particular, criteria were more likely to be considered met for female sex/gender conditions. Due to the categorical treatment of the probability that criteria were met, there was considerable uncertainty surrounding this effect. It is therefore considered in less depth here.

et al., 2012; Ratto et al., 2018; Russell et al., 2011) suggesting that greater autistic symptomology may be required if girls are to meet the diagnostic threshold.

If, as this study suggests, females require greater ASD symptom severity than males in order for diagnosticians to have the same confidence in the presence of ASD, females with less severe autistic behaviours will render diagnosticians less confident, possibly leading to a negative ASD result. This means that fewer females will fall above the 'female' severity threshold than will males (above the 'male' severity threshold). But, if a clinical diagnosis of ASD reflects individuals at the extreme end of the neurotypical distribution of ASD traits (generally higher for males than females; Lai, Lombardo, et al., 2015), the diagnostic threshold should, theoretically, be lower for females. However, this would depend on whether the surveyed diagnosticians responded about severity relative to neurotypical peers of the same sex, rather than considering non sex-specific, 'objective', functional limitations.

It is also possible that the same symptoms may appear more severe in females if these are considered to remain unmasked despite the expected camouflaging behaviours. Therefore, if diagnosticians accounted for camouflaging in their severity ratings, underlying ASD traits may have been viewed as more severe. Higher severity may not have translated to higher confidence that criteria were met as these camouflaging behaviours may have been *assumed,* but not confirmed, because they were not specifically described in the case studies.

The inconsistency between perceived symptom severity and confidence that criteria were met may also suggest that, while diagnosticians considered the presentations to be more developmentally atypical in the female sex/gender conditions, challenges inherent in assessing females may lead to less confidence in an ASD diagnosis. Given the higher prevalence of ASD in males, it is likely that diagnosticians have less experience in assessing females. Thus, it is possible that females require more severe ASD difficulties for

diagnosticians to have the same level of certainty that ASD criteria are met and overcome any 'second-guessing' or under-confidence in their impressions.

Despite almost 90% of diagnosticians reporting that ASD assessment is more challenging when the client is female, this was not mirrored in ratings of difficulty reaching a diagnosis across conditions in the female presentation case study. Instead, across both case studies, there was some evidence to suggest that assessment is perceived as more challenging when the presentation is incongruent with the sex/gender of the client. This may be because (a) the presentation is inconsistent with expectations of how ASD often manifests in clients of a given sex/gender, or (b) due to confusion as to how to quantify abnormality (i.e., relative to a typically developing female or autistic male). Regardless of the mechanism, this finding suggests that it may be difficult for diagnosticians to assess males presenting with a 'female' presentation and that even females presenting with more classic ASD may be difficult to diagnose (also noted in a study of teacher impressions; Whitlock et al., 2020). Therefore, any deviation from a male client presenting with a male ASD presentation may render assessment more challenging.

The large degree of variability in diagnosticians' ratings of difficulty in reaching a diagnosis, both between and within conditions, was also found in their confidence in the presence of ASD. This variability was unexpected, and it may be that standardised psychometric instruments (the results of which were not provided in the case studies) are useful in validating clinical impressions and thus increasing confidence in a diagnosis. Given that instruments may vary in sensitivity to the female presentation (Beggiato et al., 2017), designing instruments for use with females, or adding female norms to existing instruments, is important.

There was also evidence of bias in which diagnoses were recommended besides ASD and in the diagnoses that diagnosticians believed warranted further consideration. Attention-

deficit/hyperactivity disorder (ADHD) and the *No diagnosis* label were disproportionately ascribed in the male sex/gender conditions. Similarly, diagnosticians were more likely to recommend generalised anxiety disorder for further consideration for males than for females. A possible explanation for this is that symptoms of anxiety may be normalised more for females and thus appear more abnormal amongst males. On the other hand, the higher prevalence of ADHD among males and closer alignment of its symptoms with normative male behaviour (i.e., externalising behaviour and inattention) may explain why it was more likely to be selected for males. In sum, these findings suggest that diagnosticians' perceptions of assessment difficulty, ASD symptom severity, confidence that criteria are met, and importance of specific differential diagnoses may be differentially influenced by the sex/gender of the child assessed.

Challenges in Assessing Females for ASD

Diagnosticians reported various challenges associated with assessing females for ASD and reasons for underdiagnosis, with the majority of these pertaining to the broad theme of sex/gender differences in ASD presentation. All diagnosticians reported that males and females with ASD present differently, with almost half describing these differences as *marked*. However, diagnosticians acknowledged that not all professionals subscribe to the notion of a distinctive female ASD phenotype, and that some maintain a more traditional view of ASD (consistent with a typical 'male' presentation), perhaps regarding individuals with a 'female' phenotype as not being autistic. This suggestion of general, but not total, consensus in the existence of a female ASD phenotype is consistent with the findings of Jamison et al. (2018), in which 70% of the surveyed clinicians reported sex/gender differences in core ASD symptoms. Of the phenotypic differences identified by diagnosticians in Jamieson and colleagues' study, the majority reflected differences in RRBI behaviour severity, whereas in the present investigation, differences in both severity and

manifestation were reported across all criteria and in behaviours associated with ASD. It is likely that these were elicited through the collection of open-ended, qualitative responses. In particular, diagnosticians surveyed here placed emphasis on a "more subtle" or functional and less overt manifestation of ASD among females. Consistent with previous studies which have directly examined sex/gender differences in ASD (for a review, see Young et al., 2018), diagnosticians described less developmentally atypical restricted interests, a lack of overt difficulty in imaginative play or social interest and generally more subtle repetitive behaviour among females. Diagnosticians also reported that the female presentation of ASD may be influenced by behaviours such as camouflaging and less externalising behaviour. Broadly, diagnosticians perceived the mismatch between males and females in the manifestation of their autistic difficulties to be among the most important reasons that females may be more difficult to assess and otherwise overlooked.

Particular concern was raised regarding perceived mismatch between the female expression of ASD and the instruments, criteria, and conceptualisation by which ASD is defined. Indeed, some diagnosticians suggested that the conceptualisation of ASD for females remains unclear. Clinicians interviewed by Muggleton et al. (2019) provided some support for this, describing ASD itself as sex/gender-neutral and centred around social communication difficulties, but with sex/gender differences in social motivation, emotional recognition, and internalisation of distress affecting the outward expression of the condition. In the present study, diagnosticians' self-reported familiarity with sex/gender differences in ASD presentation varied and correlated moderately with the length of experience conducting developmental assessments. This was mirrored in some diagnosticians' reflections that familiarity with the female ASD presentation is essential for minimising false negative assessment results.

Diagnosticians expressed concern about frequent misdiagnosis and mislabelling as potential reasons for the under-detection of ASD in females. The results of the present study suggest that anxiety disorders may be particularly difficult to differentiate from ASD in females, and that there may therefore be overlap between diagnosticians' conceptualisations of these conditions. Indeed, it is possible that misdiagnosis and mislabelling serve to heighten symptoms of anxiety in females with unidentified ASD, thus perpetuating clinicians' identification of anxiety as the most salient, or indeed only, concern. This is consistent with literature highlighting elevated symptoms of anxiety in females with ASD compared to their male counterparts (Nasca et al., 2019). As a result of this symptom overlap, clear differential guidelines between anxiety disorders and female ASD may be necessary, in addition to better defining the female ASD phenotype.

Camouflaging behaviours have been associated with anxiety and other internalising concerns, such as depression and suicidal ideation (see Beck et al., 2020; Cage & Troxell-Witman, 2019; Cassidy et al., 2019) and were strongly endorsed by diagnosticians surveyed here as an important element of the female ASD presentation. There has been recent argument that camouflaging and anxiety disorders, particularly social anxiety disorder, may share an underlying conceptual basis (due to correlations of similar strength between camouflaging and social anxiety disorder, and camouflaging and ASD), suggesting discriminant validity must be established (Fombonne, 2020). Therefore, in addition to difficulty detecting camouflaging, quantifying the genuine severity of ASD difficulties and differentiating anxiety related to autistic camouflaging from anxiety without ASD may contribute to diagnosticians' lower confidence in assessing females. Camouflaging, perhaps contributing to a history of misdiagnosis and mislabelling, may also render assessment more challenging for adult women than girls or males of any age. Specifically, women who have remained undiagnosed until adulthood may be generally able to manage and/or camouflage

their ASD to a greater extent (see also Bargiela et al., 2016; Gould & Ashton-Smith, 2011), more likely to have pre-diagnosed psychiatric conditions, or simply present with more subtle ASD features than those diagnosed in childhood. Longitudinal studies of camouflaging behaviours, overt ASD difficulties, internalising difficulties and their role in the timing of diagnosis may therefore direct future research.

Circumventing Challenges

Previous findings have suggested that females may mask ASD difficulties in certain environments and that psychometric instruments may lack sensitivity to their presentation (Hull et al., 2020; Lai, Lombardo, et al., 2015). It may therefore be anticipated that diagnosticians would account for these limitations when formulating a diagnostic opinion. Indeed, the present findings did suggest that diagnosticians rely more upon parent report, and less on teacher report information, for girls than for boys. This was supported by the qualitative data suggesting that girls may effectively camouflage ASD traits at school, resulting in inconsistency between parent and teacher report (Hiller et al., 2014; Posserud et al., 2006). Further, a number of diagnosticians reported lacking trust in psychometric instruments used in quantifying females' autistic symptomology due to a perceived mismatch between their presentation and the presentation detected by the instruments. However, their estimates of their reliance on these instruments when forming a diagnostic opinion was similar for boys and girls. It is possible that, despite questioning the validity of the instruments for females, diagnosticians are inclined to use instruments when they lack confidence in a diagnosis in order to validate their impressions. This may be compounded by an absence of instruments more sensitive to the female ASD presentation. Given the above concerns raised, it is important that future research be directed to the development of psychometric instruments for the reliable detection of female ASD.

A number of practical and conceptual modifications made by diagnosticians to circumvent assessment challenges were revealed by the qualitative data, many of which related to addressing camouflaging (e.g., incorporating self-report data and collecting evidence over time). Regarding practical modifications, behavioural observation beyond the ADOS-2 (e.g., using social videos or images) and across different settings including home and school, were seen as helpful in forming a diagnostic opinion. This need to assess functional impact across a variety of settings is specified within the DSM-5 criteria (American Psychiatric Association, 2013). However, given the effects of camouflaging and differential presentations across settings, functional impact may be present, but not overtly displayed. Conceptually, therefore, an ASD diagnosis should not be ruled out based purely on an absence of obvious difficulties at school or in the assessment environment. Instead, these data suggest that a more flexible conceptualisation of ASD is needed for females: that is, consistent with males in core difficulties, but with scope for a different expression of behaviours, particularly across different settings.

Limitations

Given the broad focus of the study was communicated to participants in the recruitment material, future research could examine bias diagnostic decision-making in ASD when the importance of sex/gender is not foreshadowed. Another important limitation was, because preconstructed case studies were presented, diagnosticians were unable to formulate interview questions or use their own deductive reasoning to seek further information and form their own case conceptualisations. Consequently, it is unknown whether and how diagnostic interviewing (e.g., examples given, descriptions of abnormal behaviours) and clinical observations (i.e., the interpretation of behaviours) are influenced by a diagnostician's understanding of the female ASD presentation. Despite these limitations, this study recruited a large sample of diagnosticians and was able to triangulate quantitative and

qualitative data. It also allowed for the direct examination of sex/gender expectancy bias in ASD diagnosis, hitherto absent in the literature.

Conclusions

In this study, I have identified a large number of challenges which currently exist for diagnosticians assessing female clients for ASD. Some evidence for a sex/gender expectancy bias was found in perceived severity of difficulties (females rated as showing greater difficulty) which suggests a single ASD threshold may be inappropriate. The relationship of this difference in severity ratings and the probability that criteria were considered met was uncertain. However, there was strong evidence that higher severity ratings for female sex/gender conditions did not generalise to diagnosticians' confidence that criteria were met (or in an ASD diagnosis), or suggested support levels. It is possible that females were rated as having greater difficulty than males due to differences in neurotypical development, but that clinicians were less certain that criteria were met if the case study described a female due to implicit sex/gender differences in severity thresholds, accounting for camouflaging or second-guessing.

Diagnosticians reported that ASD assessment is more challenging for female than male clients and identified qualitative and quantitative differences in presentations between males and females as a primary reason for this. In particular, diagnosticians discussed mismatch between the female presentation and (a) stereotypical impressions of ASD among referring parties and some practitioners, and (b) diagnostic instruments used. Diagnosticians identified a number of strategies to overcome such challenges, including weighting sources of information differently and incorporating self-report and more comprehensive behavioural observation. However, there remains a need to further clarify the difficulties of females with ASD in order to develop a clearer conceptualisation and improve its identification.

Chapter 6: General Discussion

Overview

In this thesis, I examined sex/gender differences in the behavioural presentations of autistic children as a possible explanation for the asymmetry in diagnostic rates and under-identification and underdiagnosis of ASD in females. The studies presented in this thesis were designed to address the less well understood aspects of a number of broad and interrelated themes relating to sex/gender differences in the presentation of ASD. These themes are discussed below and can be categorised as pertaining to (a) how ASD symptoms are expressed, and/or (b) how symptoms are perceived, although there is considerable overlap between these categories. First, regarding the issue of ASD symptom expression, this thesis has identified several behaviours that may present differently, either qualitatively or quantitatively, among females compared to males. Sex/gender differences in symptom management strategies such as camouflaging and masking, and how distress is expressed (e.g., either internalised or externalised), were frequently raised by diagnosticians and likely moderate these phenotypic differences across different environments. In turn, camouflaging and internalisation of distress likely reflect underlying sex/gender differences in executive functioning (such as imitation) or socialised behaviours (such as desire for social interaction), both identified in this thesis. Further complicating sex/gender differences in cognitive or behavioural presentation, certain ASD characteristics may emerge later in the developmental period for females, perhaps contributing to the delay in detection of their ASD. As argued elsewhere in this thesis (Study 1), instruments designed based on androcentric literature and reflecting the male ASD presentation may lack sensitivity to the difficulties of females in the areas in which these difficulties differ. As a result of this insensitivity, instruments may underestimate the ASD difficulties of females, and for those with more subtle features, may result in a negative ASD result. Furthermore, if females are better able to conceal their

difficulties in a social situation, diagnosticians may observe less overt difficulty during assessment, either through structured observation (i.e., via instruments) or unstructured observation, and may therefore underestimate genuine ASD difficulties. In addition to diagnostician observations, teacher-reported data is also important in establishing the pervasiveness of difficulties. Therefore, if ASD behaviours are less overt in the school environment, teachers may raise less concern (as identified in this thesis), possibly resulting in delayed referral or a lack of corroborating evidence at the time of assessment.

The above issues relate to the second theme concerning the perception of ASD features, because although females may present with different ASD features, the interpretation of these features may be influenced by androcentric impressions of how ASD presents as well as biases associated with gender expectations. Moreover, referrers and diagnosticians may be primed to consider different diagnoses for a male or female, despite an identical presentation. Collectively, issues concerning the perception of ASD features among females may mean that greater ASD difficulty is required for diagnosis. Together with differences in the expression of ASD symptoms, issues relating to the perception of ASD behaviours may result in under-detection of ASD among females.

In this thesis, fine-grained sex/gender differences were examined through item-level behavioural profiles according to two widely used ASD instruments (Study 1) and in the specific difficulties that contribute to ASD criteria being met (or not met; Study 2). The inclusion of children diagnosed at a subsequent assessment after an initial negative result (Study 2a), and those with many ASD traits who had been assessed for but not diagnosed with ASD (Study 2b), allowed for novel insights into the female ASD presentation and its development over time. Diagnosticians' experiences assessing females for ASD and sex/gender bias were also explored through a mixed methods study which included an experiment consisting of case studies in which the sex/gender of the child was randomly

assigned. Broadly, the results of this thesis support and extend previous findings surrounding the existence of a distinct female presentation, which may not be widely understood, and which may be poorly captured by diagnostic instruments.

The most consistent sex/gender differences across the three studies were found in the restricted interest domain, with females rated as demonstrating less difficulty and qualitatively different specific interests compared to males. Evidence of sex/gender specific profiles of stereotypical behaviours was also found and discussed in detail below. Further, females were generally less likely to present with atypical play or an absence of social interest than males, but females may present with greater anxiety. Differences were not consistent between report sources, with females presenting with fewer overt diagnostician-observed difficulties than males at the time of assessment. This may render observation schedules insensitive to all signs and symptoms. Similarly, concern was less likely to be raised by teachers for females than males. This may be a reflection of camouflaging and other differences in the management of symptoms (e.g., internalising or externalising behaviours) which may contribute to the lack of overt difficulties in social environments and may be facilitated by females' relative skill in imitation and general social motivation.

Importantly, evidence from Study 2b suggested that sex/gender influenced how ASD-related behaviours were interpreted by diagnosticians, with some associations between behaviours and assessment outcome emerging as stronger for males, and some for females. Further, females may require more severe ASD difficulties in order for diagnosticians to be equally confident in an ASD diagnosis, suggesting that separate thresholds of clinical significance are used for males and females. Taken together, findings from this thesis may assist in better understanding the female presentation of ASD and provide further clarity as to why it may be under-detected.

Sex/Gender Differences in ASD Presentation

Sex/gender differences in the phenotypic presentation of ASD may assist in explaining why some females may not be referred for specialist assessment or meet ASD criteria if they are referred. Previous research has suggested that sex/gender differences in ASD symptom severity are small. However, this is likely due to these differences being based on females whose ASD symptom severity *exceeds* the minimum threshold of clinical significance or functional disruption and thus qualify for an ASD diagnosis. Therefore, these findings provide a minimum estimate for potential sex/gender differences, as those whose presentation differs will not be diagnosed or included in studies to date. Consistent with this line of reasoning, Lai et al. (2015) suggest that sex/gender differences in ASD may be least apparent at the level of broad diagnostic domains, as these reflect defining ASD difficulties that are essential for diagnosis. However, as an individual may meet an ASD criterion by any combination of the various difficulties related to the criterion (or behaviours which, although not obviously autistic, fit under the ASD criteria if their function is examined carefully), sex/gender differences may be most apparent at the level of the specific difficulties. Examination of fine-grained or qualitative sex/gender differences in ASD presentation were therefore a central focus of this thesis.

The largest and most consistent sex/gender differences in ASD presentation were found in the domain of restricted or obsessive interests. In Study 1, males were found to have higher scores than females on Gilliam Autism Rating Scale, 3rd Edition (GARS-3) items pertaining to superior knowledge in particular subjects (Item 48; $d = 0.35$, $HDI_{80\%} = [0.18, 0.50]$) and intense, obsessive interest in specific subjects (Item 50; $d = 0.32$, $HDI_{80\%} = [0.16, 0.48]$). Males were also 12% more likely than females to present with any atypicality on these items (i.e., to be rated at ≥ 2, the behaviour is *Somewhat like the individual*). Similarly, in Study 2b, females were found to be less likely than their male counterparts to meet DSM-5

Criterion B3 (American Psychiatric Association, 2013), but males and females who received ASD diagnoses had similar probabilities of demonstrating any atypicality in this domain. This suggests that many females showed some difficulty with restricted interests, but not enough to fully meet the criterion. Together, these findings are consistent with broader literature suggesting that restricted interests may be less pronounced among females (Allely, 2019; McFayden et al., 2019), but also suggest that females may present with difficulties that, for one reason or another, do not reach the diagnostic threshold for severity.

It is possible that restricted interests are considered sub-clinical as a result of difficulties being genuinely less pervasive or disruptive (or better disguised) among females compared to males with ASD features (with or without diagnosis), or at least perceived this way. Alternatively, or perhaps in addition, the restricted interests of females may deviate in orientation from what is typically expected of someone with ASD (i.e., they may be *atypical* or uncommon among people diagnosed with ASD). The findings presented in this thesis support both of the above possibilities. Specifically, in Study 2b, restricted interests were found to differ by sex/gender; males were more likely to present with interests in screens or vehicles, while females were more likely to present with restricted interests in particular people and craft activities, such as art and drawing. These results mirror previous findings (e.g., Hiller et al., 2014; McFayden et al., 2018) and were supported by diagnosticians surveyed in Study 3. These diagnosticians reported differences in the foci of restricted interests between males and females and cautioned that interests which manifest more socially (e.g., interests in particular people or social activities), may not be detected. Interests which manifest more socially may contribute to under-detection because these interests may appear more gender and developmentally appropriate and thus, are not considered sufficiently unusual to cause concern or perhaps to warrant mention during assessment. Additionally, certain interests, particularly those of females, may be more difficult to identify

as obsessive or disruptive if they manifest in a less intrusive or more covert manner. For example, a restricted interest in drawing (included in Study 2b under the *Craft* variable) may be overlooked as 'doodling' and although obsessive, may not necessarily be disruptive in day-to-day life and appear functional and age appropriate. On the other hand, obsessions with trains may be more problematic, requiring outings to visit the train station or the collection of timetables, etcetera. Therefore, while females may present with less difficulty related to restricted interests, it remains possible that their difficulties are less overt than those typically presented by males.

While the pervasiveness and disruptiveness of restricted interests were not examined, an important contribution of Study 2 was examination of the extent to which specific interests were predicted by sex/gender and diagnostic outcome. A novel finding of this thesis was that particular restricted interests (e.g., in random objects or vehicles) may be more strongly predictive of ASD diagnosis for males, but others (e.g., toys, animals, and specific programs/characters) may be more strongly predictive of ASD diagnosis for females. A possible explanation for the asymmetry in how strongly a restricted interest predicted ASD diagnosis is that diagnosticians may consider particular interests more atypical for males or females. There are two possible interpretations of this asymmetry in association strength. First, it is possible that diagnosticians are sensitive to the differences in the focus of restricted interests of autistic females compared to males, and thus interests are considered more unusual for one sex/gender. Alternatively, assessment protocols may be most concerned with restricted interests that are typical of ASD and may more closely reflect the male phenotype. This is concerning given diagnosticians' reliance upon these 'gold-standard' assessment instruments.

Diagnosticians who are less aware of restricted interests with which females present may not ask questions specific to these interests, resulting in the interests being overlooked.

In order to avoid overlooking restricted interests that are less common within the autistic community and ensure diagnostic interview questions are sensitive to these interests, Kreiser and White (2014) suggest first exploring any consequences of removing or interrupting engagement with an interest, or broader impacts of the interest on academic, social, or other activities, rather than placing undue emphasis on the focus of the interest.

As might be expected based on previous findings, females were also slightly less likely than males to present with stereotypical behaviours (Criterion B1). Although sex/gender differences were identified across body use, speech and language use and object use, Studies 1 and 2b provided evidence that some speech and language mannerisms may be more common among males and others more so among females. Specifically, while males had a higher probability of speech/language mannerisms broadly, and repetitive speech and echolalia in particular, females were more likely to present with use of neologisms, accents, and talking to oneself. These findings demonstrate that sex/gender specific stereotypical behavioural profiles may exist, suggesting that autistic females may not present with *fewer* but *different* stereotypical behaviours (see also Antezana et al., 2018). The implications of sex/gender specific stereotypical behaviour profiles are similar to those discussed above in the context of restricted interests. In particular, an androcentric conceptualisation of ASD is unlikely to be optimally sensitive to stereotypical mannerisms that may be present in a female client.

Findings regarding sex/differences in other RRBI features were less consistent across my studies, which is mirrored in the broader literature where findings surrounding sensory sex/gender differences remain mixed (Bitsika et al., 2018). In Study 2b, females were slightly more likely to meet Criterion B4 than males and there was evidence that some parent-reported sensory hyper-sensitivities, particularly sensory avoiding behaviours, were more likely to be reported for females. Contrary to the results of Study 2b, there was no evidence

of sex/gender differences in sensory behaviours in Study 1 and sex/gender differences in sensory features were not commonly reported by diagnosticians in Study 3. Examination of the role of client characteristics other than sex/gender, such as age or ASD severity, may assist in clarifying any sensory differences between males and females.

Two consistent sex/gender differences were identified within the broad domain of social communication. First, females whose presentations were examined in Studies 1 and 2b were found to have less difficulty than males with (or less likely to have concern raised about) imaginative or spontaneous play (see also Beggiato et al., 2017; Hiller et al., 2014). Therefore, ASD should not be excluded as a possibility if a girl's imaginative play appears typical. Second, females were more likely to be viewed by parents as behaving either too submissively or 'bossily' in play than males. The finding may relate to anecdotal suggestions that, while many autistic females may desire social interaction, the quality of these interactions may be inappropriate (Attwood et al., 2006; Holliday-Willey, 2015).

Compared to the behaviours identified above, sex/gender differences were generally less consistent in other aspects of social communication. In particular, males with ASD were found to have more difficulty with nonverbal communication than females according to scores on the CARS2-ST, and in specific nonverbal behaviours on the GARS-3. Consistent with this, females were also perceived as being less severely affected with regard to their nonverbal communication by diagnosticians in Study 3, and indeed, this was the second most frequent feature reported to differ by sex/gender. In contrast, parent-reported difficulties in nonverbal communication were more commonly documented for females than males in Study 2b. A possible reason for these differences across studies was the inclusion of subclinical, non-ASD children in Study 2b, within whom sex/gender differences in behaviours such as difficulty with eye contact were more pronounced. Other potential reasons for these differences across studies (e.g., limitations of diagnostic instruments, and differences in ASD

presentation across environments and the source of diagnostic information) are presented later in this chapter.

Although generally considered a feature associated with ASD, rather than a core characteristic of the condition, anxiety emerged as a potentially important element of the female ASD presentation. Specifically, there was probable evidence of elevated *Fear or nervousness,* or atypicality in the degree or context of anxious responses, among females compared to males in the analysis of CARS2-ST items in Study 1 (Item 10). This was the only CARS2-ST item in which there was evidence, albeit moderate, where difficulties of females exceeded those of males. This finding was consistent with the perceptions of the surveyed diagnosticians (Study 3) and also with much of the broader literature (Oswald et al., 2015; Solomon et al., 2012). Given these previous findings, stronger evidence of greater anxiety or internalising difficulties among females might have been expected. However, it is possible that, given Item 10 on the CARS2-ST measures deviation from normality in fear and nervousness (i.e., under or over fearfulness), the over fearfulness expected from females may have been diluted by consideration of the other extreme.

It is important to consider anxiety levels in neurotypical individuals in order to understand whether higher anxiety among autistic females reflects a broader pattern in the population as has been suggested in previous research (Oswald et al., 2015). Compared to the neurotypical population, anxiety has been found to be elevated in autistic individuals (Mayes et al., 2011). An important nosological question that then arises from these ideas concerns whether elevated anxiety is a consequence of the difficulties associated with ASD (e.g., social, sensory, insistence on sameness; Wigham et al., 2015), or whether anxiety constitutes a component of the disorder, particularly for females. A recent network analysis examining the relationship between ASD and anxiety (i.e., whether anxiety is associated with or an inextricable part of autism) showed that anxiety scores were “highly peripheral to ASD

scores" (Montazeri et al., 2019, p. 2227). Further, symptoms of anxiety were "dynamically similar" between neurotypical and ASD groups, suggesting that anxiety is "not a central an inextricable part of the autism realm" (p. 2228). Therefore, while diagnosticians surveyed in Study 3 discussed the challenges associated with the overlap between the conceptualisations of anxiety and ASD among females, anxiety may be best viewed as a commonly co-occurring, rather than conceptually overlapping, condition. Irrespective of whether anxiety is a result of living with ASD or a part of the female ASD presentation, awareness of this possible manifestation of distress among females is essential.

Together, sex/gender differences in stereotypical behaviours and restricted interests, and less difficulty with imaginative play, social motivation, and imitation, but increased fear/nervousness among females, corroborate and extend previous findings that differences exist between the ASD presentations of males and females.

Limitations of Diagnostic Instruments

If there is a distinctive female presentation of ASD, the diagnostic and screening instruments constructed based on androcentric literature may therefore be skewed towards the *male* ASD presentation, and hence be insensitive to females' difficulties. Therefore, ASD symptoms of females, which may be milder, managed in different ways or displayed through different behaviours, may not reach thresholds consistent with clinical diagnosis according to these instruments. In the event that symptoms do reach the diagnostic threshold, their severity may be underestimated. Here, the perceptions of diagnosticians regarding ASD instruments and their utility in assessment of females, and evidence of insensitivity of the CARS2-ST and GARS-3 are considered. Following this, I discuss implications of using instruments such as these in assessment and research of sex/gender differences, and possible avenues to address their insensitivity to female ASD.

Concern regarding the sensitivity of diagnostic instruments to the difficulties with which females present was among the most frequent challenges identified by diagnosticians surveyed in Study 3. In turn, recognising the limitations of these instruments was identified by the diagnosticians as essential in ensuring accurate assessment. Given that diagnosticians reported mistrust in the instruments, it might be expected that they would rely less heavily on these instruments to reach a diagnostic result for females than for males. However, my data did not support this conclusion. This may have in part been because the extent to which instruments were relied upon varied significantly between diagnosticians. It is also possible that diagnosticians may turn to instruments when they lack confidence in the diagnostic conclusion, but with awareness of the limitations of the instruments. As nearly 90% of diagnosticians reported that conducting an assessment for ASD is more difficult when the client is female, a lack of trusted psychometric instruments to which diagnosticians can turn to support their impressions is concerning.

Evidence supporting the concerns of diagnosticians was found in Study 1, whereby results suggested possible insensitivity of the individual items of two common ASD instruments to the presentation of ASD in females. On the CARS2-ST form in particular there were six items for which there was strong evidence of greater difficulty among males (with another three items suggesting this more weakly). The only item that favoured females, Item 10: *Fear or nervousness*, is not included within the diagnostic criteria. On balance, therefore, the asymmetry in atypicality ratings between males and females across the CARS2-ST item scores suggest that the items may be more sensitive in detecting *male* ASD difficulties, and lack sensitivity to female ASD difficulties. It is possible that the CARS2 results outlined in Study 1 may have been influenced by diagnosticians' choice of form (i.e., ST or HF), based on the presentation of the child. However, the GARS-3 results generally mirrored those of the CARS2-ST (i.e., males scored meaningfully higher on a large number

of items, and there were no items on which females' difficulties were rated as more severe). Therefore, the results of the CARS2-ST cannot be explained as being entirely a result of the choice of form. On the CARS2-ST, the sex/gender discrepancy in many items led to males scoring higher overall, but on the GARS-3, the male skew in a large number of items did not translate to a difference in index scores. Whether this was a strength of the calculation of the scaled/index scores or an indication of the relative insensitivity of index scores to the item-by-item patterns is not possible to determine from these data. Similarly, the degree of sensitivity and specificity of the CARS2 (ST and HF) and GARS-3 items to the female presentation remains unclear due to unavailability of data for children without ASD. However, the higher scores among males on many items suggest that either (a) females presented with genuinely less severe ASD difficulties, or (b) the instruments poorly represent the difficulties of females in these areas. The latter possibility may be due to items inadequately representing how a female may present a particular behaviour (e.g., lining up shells instead of matchbox cars), or exclusion of the types of behaviours through which the same underlying ASD difficulties may manifest (e.g., picking at skin instead of head banging).

Considering the concerns raised above, an important strength of Study 2 was the collection of fine-grained data on ASD-related difficulties from assessment reports, without the restrictions of instruments as to how difficulties were operationalised. For example, items on questionnaires may exclude different manifestations of any particular difficulty: in contrast to the GARS 3 Item 18, *'Seems indifferent to [an]other person's attention',* a broader variety of inappropriate social approaches were included in the operational definition for Study 2b (e.g., approaching strangers, seeking social attention at inappropriate times). Perhaps as a result of this, behaviours were identified in which females were more likely to raise concern than males, and conversely, behaviours which appeared independent of

sex/gender were found. Thus, the operationalisation and clustering of specific behaviours together (e.g., the category *Conversation content* included specific behaviours such as monologuing, speaking tangentially or speaking very little) may have influenced the overall sex/gender differences detected across studies and indeed in the broader literature.

Given the evidence presented here suggests that diagnostic instruments may lack sensitivity to the female ASD presentation, and the broad implications of this insensitivity, there are several possible avenues that could be explored. The most rigorous path would be to develop instruments used solely to identify and quantify difficulties within the female ASD presentation. This approach is currently being trialled by some researchers in the construction and validation of the Questionnaire for Autism Spectrum Conditions (Q-ASC; Attwood et al., 2011). Another option is the modification of existing tools. This could be done by adjusting the weighting of items based on the typical difficulties of males and females (e.g., less weighting on restricted interests, but more weighting on unusual fear responses for females). However, this binary approach is limited in that it would sex/gender would exclude non-binary and gender-fluid individuals. Alternatively, instruments could be adjusted by rephrasing certain items and including sample behaviours or behavioural manifestations that are more typical of females (Allely, 2019). For example, item 50 of the GARS 3, *'Shows an intense, obsessive interest in specific intellectual subjects,'* might be more sensitive to the difficulties of females if it were phrased, *'Shows an intense, obsessive interest in specific subjects (e.g., vehicles, particular objects, animals, fictional worlds).'* The construction of separate sets of norms for males and females (as exemplified by the Social Responsiveness Scale; Constantino, 2005; 2011) is yet another possibility for improving the sensitivity of instruments to the difficulties of autistic females. Implications of separate sets of norms are discussed further below.

Diagnostic Observation

Diagnostic observation, conducted either through structured observation (i.e., using standardised observation instruments) or unstructured play or interaction, is an essential component of ASD assessment. Previous findings (e.g., Lai et al., 2011) suggest that the difficulties of females may be less overt, and therefore may be overlooked by clinicians. For this reason, observational schedules such as the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al., 2012), may be particularly insensitive to the female ASD presentation (Adamou et al., 2018; Lai et al., 2011). This hypothesis was supported by three recent studies which found that women and adolescent girls diagnosed with ASD based on interview data scored significantly lower than their male counterparts based on observation through the ADOS, with some falling below the ADOS diagnostic cut-off (Adamou et al., 2018; Lai et al., 2011; Rynkiewicz & Łucka, 2015). Consistent with this, diagnosticians surveyed in Study 3 frequently reported that the ADOS may lack sensitivity to all signs of ASD in females, and that alternate and less structured diagnostic observation strategies (e.g., imaginative play or social videos/images) may better elucidate diagnostic information.

The suggestion that autistic females may demonstrate less overt difficulty than males during assessment was supported by the findings of Study 2b. Specifically, diagnosticians were slightly more likely to report concern in social communication for males than females as a result of their observations. This sex/gender difference was larger for RRBI behaviours, across both ASD and non-ASD groups. Indeed, there was only one specific behaviour for which diagnosticians were more likely to raise concern for females than for males (i.e., abnormal facial expression). This was despite considerable parent concern and indeed a higher probability of concern being raised by parents of females in some areas. Together, these findings suggest that females may present with fewer overt difficulties during

assessment than males, rendering it more challenging to identify these behaviours in either a structured or unstructured manner.

Despite this difficulty, diagnosticians reported relying on their clinical observations to a similar extent for males and females (d = 0.10, $HDI_{80\%}$ = [-0.38, 0.20]). This may be an issue for accurate diagnosis, given that females may be less likely to demonstrate observable difficulties during assessment and thus a diagnostician has fewer observations with which to triangulate reporting of difficulties from other sources. Equal reliance on clinical observations may be an issue, if what diagnosticians refer to as clinical observation in fact denotes an autistic *feel* that may be more difficult to define for females. Specifically, this autistic *feel* may differ between males and females, given differences in the extent to which difficulties are overt at the time of assessment. Females may lack the typical male autistic *feel* that has come to be associated with the disorder, and therefore clinical observations and education should be extended to ensure that diagnosticians develop an analogous *feel* for females, which may include detection of compensatory and masking strategies. It is possible that females may have their own autistic *feel,* but it is probable that this may not be as easy to detect and may currently elude even the most experienced diagnosticians.

Social Camouflaging

Compensatory and masking strategies (i.e., social camouflaging) may be disproportionality common among females and are likely important in minimising the difficulties that may be observed and thus contributing to sex/gender differences in ASD presentation. Although it was first proposed as an important element of the female ASD presentation decades ago (see Wing, 1981), camouflaging has received little empirical attention until recently. Direct investigation of camouflaging behaviour was beyond the scope of this thesis, but evidence was found to support the notion of camouflaging (or at minimum, perceived camouflaging) as a common strategy for managing ASD characteristics,

particularly for females. Consistent with the hypotheses of other authors (e.g., Attwood et al., 2006; Hull et al., 2017), camouflaging was among the most common factors identified by diagnosticians contributing to complexity in assessing females for ASD. Indeed, all diagnosticians corroborated the statement that females may be better able to camouflage their ASD difficulties, with over half of the sample strongly agreeing with this statement. Strategies aimed at compensating for, and consequently concealing ASD characteristics, were also identified by diagnosticians as primary reasons for the underdiagnosis of ASD among females, as social difficulties may be less identifiable, thus decreasing the likelihood of referral to specialist services.

If camouflaging does occur, my data suggest that more extensive camouflaging by autistic females compared to males may be the result of either or both of (a) greater social motivation (i.e., desire to 'fit in' with peers), and (b) superior ability to successfully engage in behaviours which comprise camouflaging (see also Hull et al., 2020; Tierney et al., 2016). In relation to the former, the results of Study 1 support previous findings that autistic females may present with greater desire to befriend or assimilate with peers than their male counterparts (e.g., Head et al., 2014), which has been identified by autistic females as an important motivator for camouflaging (Hull et al., 2017; Tierney et al., 2016). Specifically, this was observed in parent responses on the GARS-3, which suggested that females may be more likely than males to show a desire to make friends or interest in other people and less likely to show any atypicality in these areas. Consistent with this finding, diagnosticians generally supported the statement that females present with greater motivation to form and maintain friendships than males, and due to its relationship with camouflaging, social motivation was flagged as a reason that females' social difficulties may be overlooked (Study 3). In contrast, no evidence of a sex/gender difference in social motivation was found in Study 2. This may have been because in Study 2, this variable was operationalised to include

any atypicality (i.e., either excessive or absent social interest), and given that some females may present with a desperation to fit in with peers (Cook et al., 2017), inclusion of both extremes may have concealed any underlying differences. It remains unclear as to whether increased social motivation among females is the result of less severe, absent, or qualitatively distinct difficulty in this area, or whether this is the result of gendered socialisation, by which the importance of social behaviour is reinforced by adults (Kreiser & White, 2014).

In addition to an apparent interest in peers, superior imitation abilities among females may facilitate camouflaging. The results of pairwise comparisons in the CARS2-ST profiles in Study 1 provided strong support that imitation difficulties may be less severe among females compared to males, and illustrated that females were 15% more likely than males to present without significant difficulty in imitation. Relative skill in imitation among females is consistent with previous findings (Backer van Ommeren et al., 2017) and may be particularly important in executing compensation behaviours (aimed at bridging social communication gaps with typically developing peers; Hull et al., 2017). Further examination of the skills required for effective camouflaging may direct future investigations, particularly if these skills are considered in the light of sex/gender differences in ASD presentation. My results have emphasised the importance of future examination of the role of camouflaging in the behavioural manifestation of ASD difficulties and the relationship between camouflaging and ASD diagnosis.

The Role of Teacher-Reported Concern

The extent to which ASD-related behaviours are noticed and raised by teachers is important both in initiating a referral for specialist assessment and providing corroborating evidence of the pervasiveness of autistic behaviours at the time of assessment. Therefore, an important contribution of Study 2 and 3 was examination of the role of teacher-reported concern in the diagnostic process and outcome, hitherto largely absent in the literature. In

Study 2b, teachers were less likely to report concern for females than males across both social communication and RRBI behaviours, irrespective of diagnostic outcome. Females diagnosed with ASD were rated by teachers as showing significantly fewer difficulties than diagnosed males, and interestingly, proportions of behaviours with concern reported were similar between females who were diagnosed with ASD and males not diagnosed with ASD. With the exception of academic achievement, there were no specific areas in which teachers were more likely to report concern for females than males, and it was only for a minority of behaviours where there was no reported sex/gender difference in the likelihood of teacher concern (e.g., friendship maintenance, sensory behaviours). Together, these results are consistent with previous findings highlighting sex/gender differences in the types of concerns reported by teachers (see also Hiller et al., 2014; Mandy et al., 2011), and that teachers may report more concern for males referred for assessment than for females (Posserud et al., 2006). The probability of referral for specialist assessment may be reduced if concern is not shared by all caregivers. Further, in the event that difficulties are not reported as being pervasive across different environments, because this is diagnostic requirement, ASD diagnosis may not be provided at the time of assessment.

An important contribution of this thesis related to the above was the examination of the extent to which particular behaviours and the source of reporting were associated with ASD assessment result. In conjunction with teacher concern being generally less evident for females, the strength of the associations between teacher-reported concern and ASD diagnosis varied from strong (e.g., *Nonverbal understanding*) to weak (e.g., *Social approach*), and the number of behaviours weakly associated with assessment result was greatest for teacher report. This is consistent with the findings of Study 3, in which diagnosticians reported relying less strongly on teacher report than parent report or their own observations of the child's behaviour; attributing only 20% of their diagnostic formulation to

teacher report. Notably, there was probable evidence that diagnosticians relied less upon teacher report information for girls than for boys (d = 0.32, $HDI_{80\%}$ = [0.02, 0.62], $P_{(meaningful)}$ = 83.1%). It is unknown why this might be the case, but given the results of Studies 2b and 3, it is possible that teachers report significantly fewer concerns for girls than boys and that diagnosticians are aware of issues that may render ASD-related behaviours less overt in the school environment, or at least less likely to cause concern. Yet another possibility, supported by the findings of Study 3, is that diagnosticians are aware that the common perception of ASD among teachers and the wider community may be consistent with, and limited to, the androcentric presentation.

Lower reliance on teacher report for females compared to males may also relate to previous findings that autistic females may express their distress differently; i.e., less commonly externalised through aggression or impulsivity, and more commonly internalised through anxiety or withdrawal (Hull et al., 2016). The vast majority of diagnosticians recruited in Study 3 identified greater internalising and less externalising behaviour as features of the female presentation of ASD. Indeed, some suggested that internalising difficulties may be less overt or disruptive than externalising behaviours in the school environment. The results of Study 1 appeared to contradict those presented above, in that no meaningful sex/gender differences were observed in GARS-3 items relating to emotional responses. However, this was thought to be due to insensitivity to the *type* of emotional response (i.e., internalising rather than externalising) and inability to distinguish between these responses. The different expressions of autistic distress may not be well understood in the community, particularly those that are channelled inward. Together, these factors may allow autistic females to *fly under the radar* and result in delayed referral and/or limited teacher reported concern resulting in fewer referrals and lower probability of ASD diagnosis at the time of formal assessment.

Closely related to the finding that the degree of concern may vary according to the environment and source of reporting, diagnosticians supported the statement that, compared to males, females may present with larger differences in the presentations between the school and home environments. Differences in the overt ASD presentations of females may explain discrepancies in sex/gender differences reported across sources in Study 2b, and such discrepancies were frequently flagged by diagnosticians as a challenge during assessment. This is because an individual's difficulties must be pervasive (i.e., present across different environments) in order to qualify for an ASD diagnosis. The DSM-5 specifies that symptoms must cause *clinically significant* impairment in social, occupational, or other important areas of current functioning (American Psychiatric Association, 2013). While the DSM-5 acknowledges that symptoms may be masked, the diagnostician must use their clinical judgement to determine whether social difficulties are clinically significant and sufficiently disruptive if camouflaging is being performed convincingly. Further consideration of camouflaging and its implications may be useful in determining this. In particular, growing literature suggests that individuals may only be able to engage in camouflaging for a limited period of time, and therefore camouflaging behaviours may be contained to social environments. Further, as a result of the cognitive and emotional effort required for camouflaging coupled with its perceived necessity for fitting in with peers, social interactions, while they may appear unproblematic, may be limited in number. Finally, the suppression of distress in social environments (Beck et al., 2020; Cassidy et al., 2019; Hull et al., 2020) may result in more challenging ASD behaviour observed in the home environment than if the child's presentation were consistent across settings. Consequently, ASD diagnosis may still be appropriate despite less overt difficulties in social or school environments, if the consequences of compensatory mechanisms are sufficiently disruptive and sufficient concern is raised by parents.

Difficulties Emerging Over Time

An important hypothesis relating to camouflaging suggests that the social difficulties of females may emerge later during childhood and 'catch up' with those of males, when social demands outweigh social skills and compensatory behaviours (Hsiao et al., 2013; Mandy et al., 2018). Thus, there may be an important developmental component to sex/gender differences in ASD presentation. In Study 2a, I presented the first investigation of changes in ASD-related behaviours of females who received ASD diagnoses at a second assessment, after an initial negative ASD result. ASD is characterised as a lifelong developmental condition which is present from birth (American Psychiatric Association, 2013). Therefore, changes in the presentation of females recruited in Study 2a were not expected to reflect a genuinely later onset of ASD, but existing difficulties becoming increasingly overt and functionally disruptive.

Mandy et al. (2018) argued that later diagnosis of ASD among females may be partially the result of their social difficulties become apparent later during development. This possibility was endorsed by diagnosticians surveyed in Study 3 and may partly explain why females assessed and included in Studies 1 and 2 were, on average, slightly older than males. In Study 2a, I identified several specific social difficulties that parents were more likely to be reported as concerning at a second assessment than at the first, notably difficulties with social approach, conversation content, eye contact, facial expression, imagination and spontaneity in play, and friendship maintenance. Atypicality in conversation content (e.g., talking tangentially and providing excessive detail), was also more likely to be observed by diagnosticians and reported by teachers at the second assessment. These findings support the broad idea of Mandy et al. (2018) and extend this hypothesis by demonstrating which social behaviours are most likely to emerge over time. Although many social difficulties increased, the most meaningful difference between the first and second presentation was in the

probability that a female would meet Criterion B2: insistence on sameness, routines, and rituals. This evidence, along with elevated insistence on sameness scores among females found by Antezana et al. (2018), suggests that future investigations of sex/gender differences in the development of Criterion B2 difficulties may be worthwhile.

Due to the importance of early and specialised allied-health support (Reichow, 2012) and the necessity of diagnosis to facilitate access to this support, early diagnosis of ASD should be sought where possible. Having said this, some children (later diagnosed with ASD) may fail to meet criteria earlier in life, as shown in Study 2a. Regardless, the results of this study suggest that females who are suspected of having ASD and who present for assessment but lack specific difficulties within Criterion A1 (socio-emotional reciprocity) and/or B2 (insistence on sameness, routines and rituals) should be followed-up to monitor the potential emergence of these difficulties. Moreover, support should be provided for individuals with sub-clinical ASD traits so that their difficulties may be optimally managed and perhaps prevented from reaching clinical levels.

While it is possible that difficulties may increase in severity with development, it is also possible that limitations of current assessment protocols mean that the specific difficulties of females may not be detected reliably. Difficulties may only be detected once (and if) their manifestation becomes more *male-like*, and/or difficulties begin presenting in additional or more disruptive ways, or across different contexts. Moreover, assessment may lack sensitivity to early compensatory behaviours which may become insufficient with increasing social and environmental pressure. Given this, future research should consider the reasons for increased difficulty for females in these areas over time and any possible moderating effect of increasing anxiety and internalising difficulties upon the emergence of these difficulties. Broadly, my findings illustrate that identifying the indicators that these

children will go onto to display issues of clinical significance should be a focus for ongoing early detection research.

Sex/Gender Influence on the Interpretation of ASD Behaviours

In addition to possible developmental effects on sex/gender differences in ASD presentation, evidence presented in this thesis, particularly in Studies 2 and 3, suggested that sex/gender may influence how ASD-related behaviours are interpreted. In Study 2b, analyses of the statistical interactions between sex/gender and assessment result suggested that some behaviours were more strongly predictive of ASD diagnosis for males (e.g., deconstruction of objects), and others were more strongly predictive of ASD diagnosis for females (e.g., parent-reported speech/language mannerisms). While meaningful interactions were found in behaviours across all sources of reporting (i.e., parent report, diagnostician observation, and teacher report), they were most common in specific parent-reported stereotypical behaviours (i.e., present in six of the 23 Criterion B1 behaviours examined, see Appendix I).

There are three possible interpretations of these interactions between sex/gender and assessment result. First, diagnosticians may have interpreted these behaviours differently depending on the sex/gender of the child. Differences in interpretation may be consequent of diagnosticians considering behaviours according to their own understanding of expected sex/gender differences in ASD presentation or neurotypical children. Second, it is possible that these behaviours exist within a broader constellation of behaviours which were more or less strongly associated with ASD diagnosis or considered more or less atypical. Finally, the interactions may result from differences in the rates of the behaviours between males and females, which may also differ according to environment. That is, the behaviour in question may not be causally implicated in the diagnostic decision but related to other behaviours which result in higher or lower probability of diagnosis. Although we cannot exclude the

latter possibilities, it remains possible that behaviours, particularly stereotypical behaviours may have been interpreted differently for males or females.

Interestingly, the stereotypical behaviours presented by females were not as strongly associated with ASD diagnostic outcome (i.e., less strongly predictive of ASD diagnosis) as those presented by males (also suggested by Gould, 2017). This may contribute to females being slightly less likely to meet Criterion B1 and supports the notion that females present with different stereotypical mannerisms that may not be captured within the androcentric ASD conceptualisation. My results suggest that assessment of stereotypical behaviours should be conducted with the understanding that some behaviours commonly associated with ASD (e.g., repetitive speech, lining up objects, deconstruction) may be less common among autistic females than males. This means that diagnosis should not be withheld simply due to absence of these behaviours, but diagnosticians should consider other possible behaviours through which core ASD features may present, particularly within this criterion.

While others have suggested that referrers and clinicians may be less likely to consider ASD as a possible diagnosis for females (Hull et al., 2020), sex/gender may also influence whether clinicians consider other diagnoses for children. This sex/gender related expectancy bias may occur where features of a condition are more common in males or females or has features associated with normative behaviours of this sex/gender (Hartung & Widiger, 1998). In Study 3, diagnosticians were presented with two case studies (a male and a female ASD presentation) and the sex/gender of the child described was randomly assigned in each. Diagnosticians were found to be more likely to suggest attention-deficit/hyperactivity disorder or generalised anxiety disorder diagnoses when male sex/gender was assigned in the case studies. Given the differential developmental trajectories of neurotypical males and females, it is important that diagnosticians consider deviations in the light of what might be expected for an individual of that sex/gender (see also Koenig &

Tsatsanis, 2005). This may be particularly important regarding ASD because traits, such as social difficulties and repetitive behaviours, have been found to be more pronounced among neurotypical males compared to females (Constantino & Todd, 2003). Therefore, specific behavioural difficulties should be quantified according to the extent to which they deviate from normative characteristics of neurotypical males or females.

The results of Study 3 suggested that sex/gender influences diagnosticians' perceptions of ASD-related atypicality, wherein diagnosticians rated the severity of difficulties as greater for females, despite no meaningful difference in their confidence that the child had ASD or that criteria were met. This suggests that females require greater symptom severity than males in order for diagnosticians to have the same confidence in the presence of ASD, consistent with previous findings which suggest that greater autistic symptomology may be required of females in order to receive an ASD diagnosis (Dworzynski et al., 2012; Ratto et al., 2018; Russell et al., 2011). Diagnosticians' perceptions of the typical ASD presentations of males or females may also colour their interpretation of presenting difficulties. Specifically, there was probable evidence that diagnosticians had greater difficulty reaching a diagnostic conclusion for a female than a male presenting with a *male* ASD presentation (i.e., classic ASD). This evidence was weak for the female ASD presentation case study, but generally, evidence tended towards greater difficulty when the allocated sex/gender was incongruent with the presentation. Together, these findings suggest that both the expectations of diagnosticians regarding how ASD usually manifests in a male or female, and how unusual the behaviours are relative to neurotypical children may be important in how symptoms are perceived.

Females with Sub-Clinical ASD

An important contribution of this thesis was the consideration of presentations of children with many autistic traits but for whom the result of their formal ASD assessment

was negative. In particular, sub-clinical females were included in order to investigate how their presentations differed from females who received ASD diagnoses, and therefore where behavioural criteria may need to be broadened, and diagnosticians should (a) expect less, or at minimum, a different level of reported difficulty and be more flexible in their conceptualisation of ASD, and/or (b) probe more carefully to elucidate any difficulties in these areas.

In support of the above, although sex/gender differences emerged in a large number of parent-reported behaviours examined in Study 2b, for some behaviours, the sex/gender difference was primarily driven by non-ASD (or subclinical) children. This was particularly the case within Criterion A1, and to a lesser extent Criterion A2. For some of these behaviours, parents of ASD and non-ASD females were approximately as likely to report difficulty, but the difference in probability was far larger for males (e.g., *Sharing interests* and *Sharing emotions*). These findings support the data from Study 3, which suggested that females may require more atypicality in order for diagnosticians to have the same confidence in an ASD diagnosis (and perhaps, therefore, in order for a diagnosis to be provided). Alternatively, they may lack the breadth of ASD difficulties that must be present for diagnosis, or present with ASD-related difficulties that are not detected by diagnostic instruments or appear sufficiently atypical. Given the above, research should further investigate whether females who present with many ASD traits but fall narrowly below the diagnostic threshold should be diagnosed with ASD via more flexibly defined criteria. The key to determining the appropriateness of an ASD diagnosis may lie in accurately quantifying the functional impact of characteristics and whether ASD informed support would be beneficial.

While one may argue that the threshold for clinically significant ASD characteristics may be sex/gender specific and establishing normative data for males and females may

improve detection of females, this approach is not ideal for several reasons. ASD is a vastly heterogenous condition and this thesis, along with other research, has illustrated that an individual's sex/gender may contribute additional heterogeneity. However, other variables such as age, cognitive ability, language skill and co-occurring conditions also meaningfully influence the expression of ASD (Fombonne, 2020). Importantly, a binary approach to sex/gender is limited and does not reflect the lived experiences of many autistic people who identify as non-binary or gender-fluid (Sala et al., 2020). Therefore, the best way forward may be to broaden our conceptualisation of ASD by considering different behavioural exemplars and manifestations across different environments and stages of development in order to capture an array of different presentations.

Summary

In this thesis, I have examined sex/gender differences in the presentation of ASD as a possible reason for the observed male prevalence among diagnosed individuals. Evidence of meaningful differences in a large number of fine-grained behaviours was found. In particular, autistic females and those with subclinical presentations were found to present with different and perhaps fewer restricted interests, and a distinct profile of stereotypical behaviours. Further, females were found to have less difficulty with imagination and spontaneity in play, social motivation, and imitation than autistic males. However, females may present with higher levels of anxiety.

It has been suggested that psychometric instruments may be insensitive to the female presentation of ASD as a result of being constructed with the classic male presentation in mind (Rivet & Matson, 2011b). Some evidence of insensitivity to females' presenting difficulties was found in the specific items of the CARS2-ST and GARS-3. In addition to limitations of instruments, results suggested that females may present with fewer overt

difficulties during the time of assessment and within the school environment, perhaps due to camouflaging and/or other differences in symptom management (e.g., internalising behaviour).

The ASD-related behaviours of females may differ vastly according to environmental context, with teachers far less likely to report concern for girls than boys. In addition, ASD characteristics may also follow sex/gender specific trajectories, with females most likely to develop difficulties such as resistance to change, routines, and rituals, and males to develop difficulties with restricted interests. Indeed, of all criteria, females may be least likely to meet Criterion B3 (restricted interests), suggesting that difficulty in this domain may not be an especially common feature of the female ASD presentation.

While males and females may demonstrate qualitatively different ASD presentations, the same symptoms may be interpreted differently depending on the sex/gender of the individual. Specifically, symptoms may be perceived as more severe among females but not result in increased confidence in ASD diagnosis, and certain behaviours may be more predictive of ASD result for males, and others for females. Furthermore, diagnostic assessment may be more challenging when the sex/gender of the presentation of the client is incongruent with their sex/gender. Together, these findings suggest that diagnosticians view certain behaviours as more or less atypical compared to neurotypical children of the same sex/gender and that their expectations may affect their clinical decision-making regarding ASD.

Collectively, these difficulties may render the female ASD presentation less identifiable and contribute to misdiagnosis or *missed* diagnosis. A priority for clinical assessment protocols and future research should therefore be the adoption of a broader, more flexible conceptualisation of ASD among females.

References

Adamou, M., Johnson, M., & Alty, B. (2018). Autism Diagnostic Observation Schedule (ADOS) scores in males and females diagnosed with autism: A naturalistic study. *Advances in Autism, 4*(2), 49-55. https://doi.org/10.1108/aia-01-2018-0003

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). John Wiley & Sons.

Allely, C. S. (2019). Exploring the female autism phenotype of repetitive behaviours and restricted interests (RBRIs): A systematic PRISMA review. *Advances in Autism, 5*(3), 171-186. https://doi.org/10.1108/AIA-09-2018-0030

Alvares, G. A., Bebbington, K., Cleary, D., Evans, K., Glasson, E. J., Maybery, M. T., Pillar, S., Uljarević, M., Varcin, K., Wray, J., & Whitehouse, A. J. O. (2020). The misnomer of 'high functioning autism': Intelligence is an imprecise predictor of functional abilities at diagnosis. *Autism, 24*(1), 221-232. https://doi.org/10.1177/1362361319852831

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). American Psychiatric Publishing.

American Psychiatric Association. (2011). Guidelines for psychological practice with lesbian, gay and bisexual clients. *American Psychologist, 67*(1), 10-42.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Andersson, G. W., Gillberg, C., & Miniscalco, C. (2013). Pre-school children with suspected autism spectrum disorders: Do girls and boys have the same profiles? *Research in Developmental Disabilities, 34*(1), 413-422. https://doi.org/10.1016/j.ridd.2012.08.025

Antezana, L., Factor, R. S., Condy, E. E., Strege, M. V., Scarpa, A., & Richey, J. A. (2018). Gender differences in restricted and repetitive behaviours and interests in youth with autism. *Autism Research*. https://doi.org/10.1002/aur.2049

Asperger, H. (1944). Die “autistischen psychopathen” im kindesalter [The “autistic psychopaths” in childhood]. *Archiv für Psychiatrie und Nervenkrankheiten, 117*(1), 76-136.

Attwood, T. (2007). *The complete guide to Asperger’s syndrome*. Jessica Kingsley Publishers.

Attwood, T., Garnett, M. S., & Rynkiewicz, A. (2011). *Questionnaire for Autism Spectrum Conditions (Q-ASC)*. [Measurement instrument]. Retrieved from: https://mindsandhearts.net/gq-asc-girls-questionnaire-for-autism-spectrum-conditions/

Attwood, T., Grandin, T., Bolick, T., Faherty, C., Iland, L., Myers, J. M., & Wroble, M. (2006). *Aspergers and girls*. Future Horizons.

Backer van Ommeren, T., Koot, H. M., Scheeren, A. M., & Begeer, S. (2017). Sex differences in the reciprocal behaviour of children with autism. *Autism, 21*(6), 795-803. https://doi.org/10.1177/1362361316669622

Baldwin, S., & Costley, D. (2016). The experiences and needs of female adults with high-functioning autism spectrum disorder. *Autism, 20*(4), 483-495. https://doi.org/10.1177/1362361315590805

Banach, R., Thompson, A., Szatmari, P., Goldberg, J., Tuff, L., Zwaigenbaum, L., & Mahoney, W. (2009). Brief report: Relationship between non-verbal IQ and gender in autism. *Journal of Autism and Developmental Disorders, 39*(1), 188-193. https://doi.org/10.1007/s10803-008-0612-4

Bargiela, S., Steward, R., & Mandy, W. (2016). The experiences of late-diagnosed women with autism spectrum conditions: An investigation of the female autism phenotype.

Journal of Autism and Developmental Disorders, 46(10), 3281-3294. https://doi.org/10.1007/s10803-016-2872-8

Barnard-Brak, L., Richman, D., & Almekdash, M. H. (2019). How many girls are we missing in ASD? An examination from a clinic- and community-based sample. *Advances in Autism, 5*(3), 214-224. https://doi.org/10.1108/AIA-11-2018-0048

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

Baron-Cohen, S., & Hammer, J. (1997). Is autism an extreme form of the male brain? *Advanced Infancy Research, 11*, 193-217.

Beck, J. S., Lundwall, R. A., Gabrielsen, T., Cox, J. C., & South, M. (2020). Looking good but feeling bad: "Camouflaging" behaviors and mental health in women with autistic traits. *Autism, 24*(4), 809-821. https://doi.org/10.1177/1362361320912147

Begeer, S., Mandell, D., Wijnker-Holmes, B., Venderbosch, S., Rem, D., Stekelenburg, F., & Koot, H. M. (2013). Sex differences in the timing of identification among children and adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(5), 1151-1156. https://doi.org/10.1007/s10803-012-1656-z

Beggiato, A., Peyre, H., Maruani, A., Scheid, I., Rastam, M., Amsellem, F., Gillberg, C. I., Leboyer, M., Bourgeron, T., Gillberg, C., & Delorme, R. (2017). Gender differences in autism spectrum disorders: Divergence among specific core symptoms. *Autism Research, 10*(4), 680-689. https://doi.org/10.1002/aur.1715

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. *British Journal of Psychiatry, 201*, 116-123. https://doi.org/10.1192/bjp.bp.111.097899

Bitsika, V., & Sharpley, C. F. (2019). Effects of diagnostic severity upon sex differences in behavioural profiles of young males and females with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-019-04159-x

Bitsika, V., Sharpley, C. F., & Mills, R. (2018). Sex differences in sensory features between boys and girls with autism spectrum disorder. *Research in Autism Spectrum Disorders, 51*, 49-55. https://doi.org/10.1016/j.rasd.2018.04.002

Blakemore, S. J., Winston, J., & Frith, U. (2004). Social cognitive neuroscience: Where are we heading? *Trends in Cognitive Sciences, 8*(5), 216-222. https://doi.org/10.1016/j.tics.2004.03.012

Bölte, S., Duketis, E., Poustka, F., & Holtmann, M. (2011). Sex differences in cognitive domains and their clinical correlates in higher-functioning autism spectrum disorders. *Autism, 15*(4), 497-511. https://doi.org/10.1177/1362361310391116

Boorse, J., Cola, M., Plate, S., Yankowitz, L., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2019). Linguistic markers of autism in girls: Evidence of a "blended phenotype" during storytelling. *Molecular Autism, 10*(14). https://doi.org/10.1186/s13229-019-0268-2

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage Publications.

Cage, E., & Burton, H. (2019). Gender differences in the first impressions of autistic adults. *Autism Research*. https://doi.org/10.1002/aur.2191

Cage, E., & Troxell-Witman, Z. (2019). Understanding the reasons, contexts and costs of camouflaging for autistic adults. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-03878-x

Carter, A. S., Black, D. O., Tewani, S., Connolly, C. E., Kadlec, M. B., & Tager-Flusberg, H. (2007). Sex differences in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(1), 86-97. https://doi.org/10.1007/s10803-006-0331-7

Cassidy, S., Bradley, L., Shaw, R., & Baron-Cohen, S. (2018). Risk markers for suicidality in autistic adults. *Molecular Autism, 9*(42). https://doi.org/10.1186/s13229-018-0226-4

Cassidy, S. A., Gould, J., Townsend, E., Pelton, M., Robertson, A. E., & Rodgers, J. (2019). Is camouflaging autistic traits associated with suicidal thoughts and behaviours? Expanding the Interpersonal Psychological Theory of Suicide in an undergraduate student sample. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-019-04323-3

Chen, M. T., Chang, Y. P., Lu, X., Simeonsson, R. J., & Marraccini, M. E. (2020). *Modeling and explaining sex differences in the prevalence of autism spectrum disorder: A meta-analytic approach*. Retrieved from https://doi.org/10.31219/osf.io/kbev3.

Cheslack-Postava, K., & Jordan-Young, R. M. (2012). Autism spectrum disorders: Toward a gendered embodiment model. *Social Science and Medicine, 74*, 1667-1674. https://doi.org/10.1016/j.socscimed.2011.06.013

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates.

Cola, M. L., Plate, S., Yankowitz, L., Petrulla, V., Bateman, L., Zampella, C. J., de Marchena, A., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2020). Sex differences in the first impressions made by girls and boys with autism. *Molecular Autism, 11*(49). https://doi.org/10.1186/s13229-020-00336-3

Constantino, J. N. (2005). *Social Responsiveness Scale (SRS)*. [Measurement instrument]. Pearson Clinical Assessment.

Constantino, J. N. (2011). *Social Responsiveness Scale, Second Edition (SRS-2)*. [Measurement instrument]. Pearson Clinical Assessment.

Constantino, J. N., & Charman, T. (2012). Gender bias, female resilience, and the sex ratio in autism. *Journal of the American Academy of Child Adolescent Psychiatry, 51*(8), 756-758. https://doi.org/10.1016/j.jaac.2012.05.017

Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population: A twin study. *Archives of General Psychiatry, 60*(5), 524-530. https://doi.org/10.1001/archpsyc.60.5.524

Cook, A., Ogden, J., & Winstone, N. (2017). Friendship motivations, challenges and the role of masking for girls with autism in contrasting school settings. *European Journal of Special Needs Education*, 1-14. https://doi.org/10.1080/08856257.2017.1312797

Cridland, E. K., Jones, S. C., Caputi, P., & Magee, C. A. (2014). Being a girl in a boys' world: Investigating the experiences of girls with autism spectrum disorders during adolescence. *Journal of Autism and Developmental Disorders, 44*, 1261-1274. https://doi.org/10.1007/s10803-013-1985-6

Cumming, G. (2012). *Understanding the New Statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge/Taylor & Francis Group.

Daniels, A. M., & Mandell, D. S. (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism, 18*(5), 583-597. https://doi.org/10.1177/1362361313480277

Denwood, M. J. (2016). runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software, 71*. https://doi.org/10.18637/jss.v071.i09

Dunn, W. (1999). *Sensory Profile*. [Measurement Instrument]. Psychological Corporation.

Duvekot, J., van der Ende, J., Verhulst, F. C., Sleppendel, G., van Daalen, E., Maras, A., & Greaves-Lord, K. (2016). Factors influencing the probability of a diagnosis of autism spectrum disorder in girls versus boys. *Autism*. https://doi.org/10.1177/1362361316672178

Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child Adolescent Psychiatry, 51*(8), 788-797. https://doi.org/10.1016/j.jaac.2012.05.018

Ferri, S. L., Abel, T., & Brodkin, E. S. (2018). Sex differences in autism spectrum disorder: A review. *Current Psychiatry Reports, 20*(9), 9. https://doi.org/10.1007/s11920-018-0874-2

Fombonne, E. (2005). Epidemiology of autistic disorder and other pervasive developmental disorders. *Journal of Clinical Psychiatry, 66*, 3-8.

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research, 65*, 591-598. https://doi.org/10.1203/PDR.0b013e31819e7203

Fombonne, E. (2020). Camouflage and autism. *Journal of Child Psychology and Psychiatry, 61*(7), 735-738. https://doi.org/10.1111/jcpp.13296

Frazier, T. W., & Hardan, A. Y. (2017). Equivalence of symptom dimensions in females and males with autism. *Autism, 21*(6), 749-459. https://doi.org/10.1177/1362361316660066

Frith, U. (1991). *Autism and Asperger Syndrome*. Cambridge University Press.

Geelhand, P., Bernard, P., Klein, O., van Tiel, B., & Kissine, M. (2019). The role of gender in the perception of autism symptom severity and future behavioural development. *Molecular Autism, 10*(16). https://doi.org/10.1186/s13229-019-0266-4

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC Texts in Statistical Science.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*(2), 189-211. https://doi.org/10.1080/19345747.2011.618213

Giarelli, E., Wiggins, L. D., Rice, C. E., Levy, S. E., Kirby, R. S., Pinto-Martin, J., & Mandell, D. (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and Health Journal, 3*(2), 107-116. https://doi.org/10.1016/j.dhjo.2009.07.001

Gilliam, J. E. (2014). *Gilliam Autism Rating Scale* (3rd ed.). [Measurement instrument]. Pro-Ed Inc.

Goldman, S. (2013). Sex, gender and the diagnosis of autism: A biosocial view of the male preponderance. *Research in Autism Spectrum Disorders, 7*(6), 675-679. https://doi.org/10.1016/j.rasd.2013.02.006

Gomot, M., & Wicker, B. (2012). A challenging, unpredictable world for people with autism spectrum disorder. *International Journal of Psychophysiology, 83*(2), 240-247. https://doi.org/10.1016/j.ijpsycho.2011.09.017

Gould, J. (2017). Towards understanding the under-recognition of girls and women on the autism spectrum. *Autism, 21*(6), 703-705.

Gould, J., & Ashton-Smith, J. (2011). Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum. *Good Autism Practice, 12*(1), 34-41.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience, 9*(10), 1218-1220.

Harrop, C., Jones, D., Zheng, S., Nowell, S. W., Boyd, B. A., & Sasson, N. (2018). Sex differences in social attention in autism spectrum disorder. *Autism Research, 11*(9), 1264-1275. https://doi.org/10.1002/aur.1997

Hartley, S. L., & Sikora, D. M. (2009). Sex differences in autism spectrum disorder: An examination of developmental functioning, autistic symptoms, and coexisting behavior problems in toddlers. *Journal of Autism and Developmental Disorders, 39*(12), 1715-1722. https://doi.org/10.1007/s10803-009-0810-8

Hartung, C. M., & Widiger, T. A. (1998). Gender differences in the diagnosis of mental disorders: Conclusions and controversies of the DSM-IV. *Psychological Bulletin, 123*(3), 260-278.

Head, A. M., McGillivray, J. A., & Stokes, M. A. (2014). Gender differences in emotionality and sociability in children with autism spectrum disorders. *Molecular Autism, 5*(19). https://doi.org/10.1186/2040-2392-5-19

Hiller, R. M., Young, R. L., & Weber, N. (2014). Sex differences in autism spectrum disorder based on DSM-5 criteria: Evidence from clinician and teacher reporting. *Journal of Abnormal Child Psychology, 42*(8), 1381-1393. https://doi.org/10.1007/s10802-014-9881-x

Holliday-Willey, L. (2015). *Pretending to be normal: Living with Aspeger's syndrome (autism spectrum disorder)*. Jessica Kingsley Publishers.

Hsiao, M. N., Tseng, W. L., Huang, H. Y., & Gau, S. S. F. (2013). Effects of autistic traits on social and school adjustment in children and adolescents: The moderating roles of age and gender. *Research in Developmental Disabilities, 34*(1), 254-265. https://doi.org/10.1016/j.ridd.2012.08.001

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Hull, L., Lai, M. C., Baron-Cohen, S., Allison, C., Smith, P., Petrides, K. V., & Mandy, W. (2019). Gender differences in self-reported camouflaging in autistic and non-autistic adults. *Autism*. https://doi.org/10.1177/1362361319864804

Hull, L., & Mandy, W. (2017). Protective effect or missed diagnosis? Females with autism spectrum disorder. *Future Neurology, 12*(3), 159-169. https://doi.org/10.2217/fnl-2017-0006

Hull, L., Mandy, W., Lai, M. C., Baron-Cohen, S., Allison, C., Smith, P., & Petrides, K. V. (2018). Development and validation of the Camouflaging Autistic Traits Questionnaire (CAT-Q). *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3792-6

Hull, L., Mandy, W., & Petrides, K. V. (2016). Behavioural and cognitive sex/gender differences in autism spectrum condition and typically developing males and females. *Autism, 21*(6), 706-727. https://doi.org/10.1177/1362361316669087

Hull, L., Petrides, K. V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M. C., & Mandy, W. (2017). "Putting on my best normal": Social camouflaging in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-017-3166-5

Hull, L., Petrides, K. V., & Mandy, W. (2020). The female autism phenotype and camouflaging: A narrative review. *Review Journal of Autism and Developmental DIsorders*. https://doi.org/10.1007/s40489-020-00197-9

Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., Beckmann, J. S., Rosenfeld, J. A., & Eichler, E. E. (2014). A higher mutational

burdern in females supports a ‘Female Protective Model’ in neurodevelopmental disorders. *The American Journal of Human Genetics, 94*, 415-425. https://doi.org/10.1016/j.ajhg.2014.02.001

Jamison, R., Bishop, S., Huerta, M., & Halladay, A. K. (2018). The clinician perspective on sex differences in autism spectrum disorders. *Autism, 21*(6), 772–784. https://doi.org/10.1177/1362361316681481

Kaat, A. J., Shui, A. M., Ghods, S. S., Farmer, C. A., Esler, A. N., Thurm, A., Georgiades, S., Kanne, S. M., Lord, C., Kim, Y. S., & Bishop, S. L. (2020). Sex differences in scores on standardized measures of autism symptoms: A multisite integrative data analysis. *The Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13242

Kanfiszer, L., Davies, F., & Collins, S. (2017). ‘I was just so different’: The experiences of women diagnosed with an autism spectrum disorder in adulthood in relation to gender and social relationships. *Autism*, 1-9. https://doi.org/10.1177/1362361316687987

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism, 20*(4), 442-462. https://doi.org/10.1177/1362361315588200

Kirkovski, M., Eniticott, P. G., & Fitzgerald, P. B. (2013). A review of the role of female gender in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*, 2584-2603. https://doi.org/10.1007/s10803-013-1811-1

Knickmeyer, R., Wheelwright, S. J., & Baron-Cohen, S. (2008). Sex-typical play: Masculinization/defeminization in girls with an autism spectrum condition. *Journal of Autism and Developmental Disorders, 38*(6), 1028-1035. https://doi.org/10.1007/s10803-007-0475-0

Koenig, K., & Tsatsanis, K. D. (2005). Pervasive developmental disorders in girls. In D. J. Bell, S. I. Foster, & E. J. Mash (Eds.), *Handbook of behavioural and emotional problems in girls* (pp. 211-237). Kluwer Academic Publishing.

Kopp, S., & Gillberg, C. (1992). Girls with social deficits and learning problems: Autism, atypical Asperger syndrome or a variant of these conditions. *European Child and Adolescent Psychiatry, 1*(2), 89-99.

Kothari, R., Skuse, D., Wakefield, J., & Micali, N. (2013). Gender differences in the relationship between social communication and emotion recognition. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(11), 1148-1157. https://doi.org/10.1016/j.jaac.2013.08.006

Kreiser, N. L., & White, S. W. (2014). ASD in females: Are we overstating the gender difference in diagnosis? *Clinical Child and Family Psychological Review, 17*(1), 67-84. https://doi.org/10.1007/s10567-013-0148-9

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14*, 293-300. https://doi.org/10.1016/j.tics.2010.05.001

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS and Stan* (2nd ed.). Elsevier Inc.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science, 1*(2), 270-280. https://doi.org/10.1177/2515245918771304

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*, 178-206. https://doi.org/10.3758/s13423-016-1221-4

Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in heirarchical models. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of compuational and mathematical psychology*. Oxford University Press.

Kumazaki, H., Muramatsu, T., Kosaka, H., Fujisawa, T. X., Iwata, K., Tomoda, A., Tsuchiya, K., & Mimura, M. (2015). Sex differences in cognitive and symptom profiles in children with high functioning autism spectrum disorders. *Research in Autism Spectrum Disorders, 13-14*, 1-7. https://doi.org/10.1016/j.rasd.2014.12.011

Lai, M. C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry, 2*(11), 1013-1027. https://doi.org/10.1016/s2215-0366(15)00277-1

Lai, M. C., Baron-Cohen, S., & Buxbaum, J. D. (2015). Understanding autism in the light of sex/gender. *Molecular Autism, 6*(24). https://doi.org/10.1186/s13229-015-0021-4

Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: Setting the scene for future research. *Journal of the American Academy of Child and Adolescent Psychiatry, 54*(1), 11-24. https://doi.org/10.1016/j.jaac.2014.10.003

Lai, M. C., Lombardo, M. V., Pasco, G., Ruigrok, A. N. V., Wheelwright, S. J., Sadek, S. A., Chakrabarti, B., Baron-Cohen, S., & MRC AIMS Consortium. (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PLoS One, 6*(6), e20835. https://doi.org/10.1371/journal.pone.0020835

Lai, M. C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., Happe, F., & Baron-Cohen, S. (2016). Quantifying and exploring camouflaging in men and women with autism. *Autism*. https://doi.org/10.1177/1362361316671012

Lai, M. C., Lombardo, M. V., Ruigrok, A. N. V., Chakrabarti, B., Wheelwright, S. J., Auyeung, B., Allison, C., Consortium, M. A., & Baron-Cohen, S. (2012). Cognition

in males and females with autism: Similarities and differences. *PLoS One, 7*(10), e47198. https://doi.org/10.1371/journal.pone.0047198

Lai, M. C., & Szatmari, P. (2019). Sex and gender impacts on the behavioural presentation and recognition of autism. *Current Opinion in Psychiatry*. https://doi.org/10.1097/YCO.0000000000000575

Lam, K. S. L., & Aman, M. G. (2007). The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(5), 855-866. https://doi.org/10.1007/s10803-006-0213-z

Lehnhardt, F. G., Falter, C. M., Gawronski, A., Pfeiffer, K., Tepest, R., Franklin, J., & Vogeley, K. (2016). Sex-related cognitive profile in autism spectrum disorders diagnosed late in life: Implications for the female autistic phenotype. *Journal of Autism and Developmental Disorders, 46*(1), 139-154. https://doi.org/10.1007/s10803-015-2558-7

Leman, P. J., & Tenenbaum, H. R. (2011). Practicing gender: Children's relationships and the development of gendered behaviour and beliefs. *The British Journal of Developmental Psychology, 29*, 153-157. https://doi.org/10.1111/j.2044-835X.2011.02032.x

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology, 79*, 328-348. https://doi.org/10.1016/j.jesp.2018.08.009

Livingston, L. A., Colvert, E., The Social Relationships Study Team, Bolton, P., & Happé, F. (2019). Good social skills despite poor theory of mind: Exploring compensation in autism spectrum disorder. *The Journal of Child Psychology and Psychiatry, 60*(1), 102-110. https://doi.org/10.1111/jcpp.12886

Lombardo, M. V., Lai, M. C., & Baron-Cohen, S. (2019). Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular Psychiatry, 24*, 1435-1450. https://doi.org/10.1038/s41380-018-0321-0

Loomes, R., Hull, L., & Mandy, W. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic meta-analysis. *Journal of the American Academy of Child Adolescent Psychiatry, 56*(6), 466-474. https://doi.org/10.1016/j.jaac.2017.03.013

Lord, C., Risi, S., Lambrecht, L., Cook, E. H. J., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1-4*. Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659-685.

Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2011). Sex differences in autism spectrum disorder: Evidence from a large sample of children and adolescents. *Journal of Autism and Developmental Disorders, 42*(7), 1304-1313. https://doi.org/10.1007/s10803-011-1356-0

Mandy, W., Pellicano, L., St Pourcain, B., Skuse, D., & Heron, J. (2018). The development of autistic social traits across childhood and adolescence in males and females.

Journal of Child Psychology and Psychiatry, 59(11), 1143-1151. https://doi.org/10.1111jcpp.12913

Matson, J. L., Matheis, M., Burns, C. O., Esposito, G., Venuti, P., Pisula, E., Misiak, A., Kalyva, E., Tsakiris, V., Kamio, Y., Ishitobi, M., & Goldin, R. L. (2017). Examining cross-cultural differences in autism spectrum disorder: A multinational comparison from Greece, Italy, Japan, Poland, and the United States. *European Psychiatry, 42*, 70-76. https://doi.org/10.1016/j.eurpsy.2016.10.007

Mayes, S. D., Calhoun, S. L., Murray, M. J., Ahuja, M., & Smith, L. (2011). Anxiety, depression, and irritability in children with autism relative to other neuropsychiatric disorders and typical development. *Research in Autism Spectrum Disorders, 5*(1), 474-485. https://doi.org/10.1016/j.rasd.2010.06.012

Mayes, S. D., & Lockridge, R. (2018). Brief report: How accurate is teacher report of autism symptoms compared to parent report? *Journal of Autism and Developmental Disorders, 48*(5), 1833-1840. https://doi.org/10.1007/s10803-017-3325-8

McFayden, T. C., Albright, J., Muskett, A. E., & Scarpa, A. (2018). Brief report: Sex differences in ASD diagnosis- A brief report on restricted interests and repetitive behaviours. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3838-9

McFayden, T. C., Antezana, L., Albright, J., Muskett, A. E., & Scarpa, A. (2019). Sex differences in an autism spectrum disorder diagnosis: Are restricted repetitive behaviours and interests the key? *Review Journal of Autism and Developmental DIsorders*. https://doi.org/10.1007/s40489-019-00183-w

McLennan, J. D., Lord, C., & Scholper, E. (1993). Sex differences in higher functioning people with autism. *Journal of Autism and Developmental Disorders, 23*(2), 217-227.

Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., Carver, L., Charman, T., Chawarska, K., Curtin, S., Dobkins, K., Hertz-Picciotto, I., Hutman, T., Iverson, J. M., Landa, R., Nelson, C. A., Stone, W. L., Tager-Flusberg, H., & Zwaigenbaum, L. (2015). Early sex differences are not autism-specific: A Baby Siblings Research Consortium (BSRC) study. *Molecular Autism, 6*(32). https://doi.org/10.1186/s13229-015-0027-y

Miodovnik, A., Harstad, E., Sideridis, G., & Huntington, N. (2015). Timing of the diagnosis of attention-deficit/hyperactivity disorder and autism spectrum disorder. *Pediatrics, 136*(4), 830-837. https://doi.org/10.1542/peds.2015-1502

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2015). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*, 103-123. https://doi.org/10.3758/s13423-015-0947-8

Muggleton, J. T. B., MacMahon, K., & Johnston, K. (2019). Exactly the same but completely different: A thematic analysis of clinical psychologists' conceptions of autism across genders. *Research in Autism Spectrum Disorders, 62*, 75-84. https://doi.org/10.1016/j.rasd.2019.03.004

Nasca, B. C., Lopata, C., Donnelly, J. P., Rodgers, J. D., & Thomeer, M. L. (2019). Sex differences in externalizing and internalizing symptoms of children with ASD. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-019-04132-8

Navot, N., Jorgenson, A. G., & Webb, S. J. (2017). Maternal experience raising girls with autism spectrum disorder: A qualitative study. *Child Care, Health and Development, 43*(4), 546-545. https://doi.org/10.1111/cch.12470

Nowell, S. W., Jones, D. R., & Harrop, C. (2019). Circumscribed interests in autism: Are there sex differences? *Advances in Autism*. https://doi.org/10.1108/AIA-09-2018-0032

Øien, R. A., Hart, L., Schjølberg, S., Wall, C. A., Kim, E. S., Nordahl-Hansen, A., Eisemann, M. R., Chawarska, K., Volkmar, F. R., & Shic, F. (2017). Parent-endorsed sex differences in toddlers with and without ASD: Utilizing the M-CHAT. *Journal of Autism and Developmental Disorders, 47*, 126-134. https://doi.org/10.1007/s10803-016-2945-8

Øien, R. A., Vambheim, S. M., Hart, L., Nordahl-Hansen, A., Erickson, C., Wink, L., Eisemann, M. R., Shic, F., Volkmar, F. R., & Grodberg, D. (2018). Sex-differences in children referred for assessment: An exploratory factor analysis of the Autism Mental Status Exam (AMSE). *Journal of Autism and Developmental Disorders, 48*, 2286-2292. https://doi.org/10.1007/s10803-018-3488-y

Oswald, T. M., Winter-Messiers, M. A., Gibson, B., Schmidt, A. M., Herr, C. M., & Solomon, M. (2015). Sex differences in internalizing problems during adolescence in austism spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(2), 624-636. https://doi.org/10.1007/s10803-015-2608-1

Petrou, A. M., Parr, J. R., & McConachie, H. (2018). Gender differences in parent-reported age at diagnosis of children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 50*, 32-42. https://doi.org/10.1016/j.rasd.2018.02.003

Pisula, E., Pudło, M., Słowińska, M., Kawa, R., Strząska, M., Banasiak, A., & Wolańczyk, T. (2017). Behavioral and emotional problems in high-functioning girls and boys with autism spectrum disorders: Parents' reports and adolescents' self-reports. *Autism, 21*(6), 738-748. https://doi.org/10.1177/1362361316675119

Plummer, M. (2017). *JAGS veresion 4.3.0 user manual*. [Computer software manual]. Retrieved from https://sourceforge.net/projects/mcmc-jags/files/

Posserud, M., Lundervold, A., & Gillberg, C. (2006). Autistic features in a total population of 7- to 9-year-old children assessed by the ASSQ (Autism Spectrum Screening

Questionnaire). *Journal of Child Psychology and Psychiatry, 47*, 167-175. https://doi.org/10.1111/j.1469-7610.2005.01462.x

Potts, M. K., Burnam, M. A., Wells, K. B., Rimal, R., Siegel, R., & Potts, D. L. (1991). Gender differences in depression detection: A comparison of clinician diagnosis and standardized assessment. *Journal of Counselling and Clinical Psychology, 3*(4), 609-615. https://doi.org/10.1037/1040-3590.3.4.609

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rabbitte, K., Prendeville, P., & Kinsella, W. (2017). Parents' experiences of the diagnostic process for girls with autism spectrum disorder in Ireland: An interpretative phenomenological analysis. *Educational and Child Psychology, 34*(2), 54-66. https://doi.org/10.1080/03033910.2014.982143

Ratto, A. B., Kenworthy, L., Yerys, B. E., Bascome, J., Trubanova Wieckowski, A., White, S. W., Wallace, G. L., Pugliese, C., Schultz, R. T., Ollendick, T. H., Scarpa, A., Seese, S., Register-Brown, K., Martin, A., & Anthony, L. G. (2018). What about the girls? Sex-based differences in autistic traits and adaptive skills. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-017-3413-9

Reichow, B. (2012). Overview of meta-analyses on early intensive behaviour intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*, 512-520. https://doi.org/10.1007/s10803-011-1218-9

Reinhardt, V. P., Wetherby, A. M., Schatschneider, C., & Lord, C. (2015). Examination of sex differences in a large sample of young children with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders, 45*(3), 697-706. https://doi.org/10.1007/s10803-014-2223-6

Rivet, T. T., & Matson, J. L. (2011a). Gender differences in core symptomatology in autism spectrum disorders across the lifespan. *Journal of Developmental and Physical Disabilities, 23*, 399-420. https://doi.org/0.1007/s10882-011-9235-3

Rivet, T. T., & Matson, J. L. (2011b). Review of gender differences in core symptomatology in autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(3), 957-976. https://doi.org/10.1016/j.rasd.2010.12.003

Robinson, E. B., Lichtenstein, P., Anckarsater, H., Happé, F., & Ronald, A. (2013). Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences, 110*(13), 5258-5262. https://doi.org/10.1073/pnas.1211070110

Robinson, E. B., St Pourcain, B., Anttila, B., Kosmicki, V., Bulik-Sullivan, J. A., Grove, B., J.,, Maller, J., Samocha, K. E., Sanders, S. J., Ripke, S., Martin, J., Hollegaard, M. V., Werge, T., Hougaard, D. M., iPSYCH-SSI-Broad Autism Group, Neale, B. M., Evans, D. M., Skuse, D., Mortensen, P. B., Børglum, A. D., Ronald, A., Smith, G. D., & Daly, M. J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature Genetics, 48*(5), 552-555. https://doi.org/10.1038/ng.3529

Russell, G., Steer, C., & Golding, J. (2011). Social and demographic factors that influence the diagnosis of autistic spectrum disorders. *Social Psychiatry and Psychiatric Epidemiology, 46*(12), 1283-1293. https://doi.org/10.1007/s00127-010-0294-z

Rutherford, M., McKenzie, K., Johnson, T., Catchpole, C., O'Hare, A., McClure, I., Forsyth, K., McCartney, D., & Murray, A. (2016). Gender ratio in a clinical population sample, age of diagnosis and duration of assessment in children and adults with autism spectrum disorder. *Autism, 20*(5), 628-634. https://doi.org/10.1177/1362361315617879

Rutter, M., Caspi, A., & Moffitt, T. E. (2003). Using sex differences in psychopathology to study causal mechanisms: Unifying issues and research strategies. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44*(8), 1092-1115. https://doi.org/10.1111/1469-7610.00194

Rynkiewicz, A., & Łucka, I. (2015). Autism spectrum disorder (ASD) in girls. Co-occurring psychopathology. Sex differences in clinical manifestation. *Psychiatria Polska, 31*. https://doi.org/10.12740/PP/OnlineFirst/58837

Rynkiewicz, A., Schuller, B., Marchi, E., Piana, S., Camurri, A., Lassalle, A., & Baron-Cohen, S. (2016). An investigation of the 'female camouflage effect' in autism using a computerized ADOS-2 and a test of sex/gender differences. *Molecular Autism, 7*, 10. https://doi.org/10.1186/s13229-016-0073-0

Sala, G., Pecora, L., Hooley, M., & Stokes, M. A. (2020). As diverse as the spectrum itself: Trends in sexuality, gender and autism. *Current Developmental Disorders Reports*. https://doi.org/10.1007/s40474-020-00190-1

Sandin, S., Lichtenstein, P., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *Journal of the American Medical Association, 311*(17), 1770-1777. https://doi.org/10.1001/jama.2014.4144

Scholper, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale (CARS)*. [Measurement Instrument]. Western Psychological Services.

Scholper, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *The Childhood Autism Rating Scale- Second Edition (CARS2)*. [Measurement Instrument]. Western Psychological Services.

Schuck, R. K., Flores, R. E., & Fung, L. K. (2019). Brief report: Sex/gender differences in symptomology and camouflaging in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-019-03998-y

Sedgewick, F., Hill, V., & Pellicano, E. (2019). 'It's different for girls': Gender differences in the friendships and conflict of autistic and neurotypical adolescents. *Autism, 23*(5), 1119-1132. https://doi.org/10.1177/1362361318794930

Sedgewick, F., Hill, V., Yates, R., Pickering, L., & Pellicano, E. (2016). Gender differences in the social motivation and friendship experiences of autistic and non-autistic adolescents. *Journal of Autism and Developmental Disorders, 46*(4), 1297-1306. https://doi.org/10.1007/s10803-015-2669-1

Siklos, S., & Kerns, K. A. (2007). Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. *Research in Developmental Disabilities, 28*(1), 9-22. https://doi.org/10.1016/j.ridd.2005.09.003

Solomon, M., Miller, M., Taylor, S. L., Hinshaw, S. P., & Carter, C. S. (2012). Autism symptoms and internalizing psychopathology in girls and boys with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*(1), 48-59. https://doi.org/10.1007/s10803-011-1215-z

Springer, K. W., Stellman, J. M., & Jordan-Young, R. M. (2011). Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health. *Social Science and Medicine, 74*. https://doi.org/10.1016/j.socscimed.2011.05.033

Szatmari, P., Liu, X. Q., Zwaigenbaum, L., Paterson, A. D., Woodbury-Smith, M., Georgiades, S., Duku, E., & Thompson, A. (2012). Sex differences in repetitive stereotyped behaviours in autism: Implications for genetic liability. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 1*, 5-12. https://doi.org/10.1002/ajmg.b.31238

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, 43*(11), 2664-2676. https://doi.org/10.1007/s10803-013-1819-6

Tierney, S., Burns, J., & Kilbey, E. (2016). Looking behind the mask: Social coping strategies of girls on the autistic spectrum. *Research in Autism Spectrum Disorders, 23*, 73-83. https://doi.org/10.1016/j.rasd.2015.11.013

Tillmann, J., Ashwood, K., Absoud, M., Bölte, S., Bonnet-Brilhault, F., Buitelaar, J. K., Calderoni, S., Calvo, R., Canal-Bedia, R., Canitano, R., De Bildt, A., Gomot, M., Hoekstra, P. J., Kaale, A., McConachie, H., Murphy, D. G., Narzisi, A., Oosterling, I., Pejovic-Milovancevic, M., Persico, A. M., Puig, O., Roeyers, H., Rommelse, N., Sacco, R., Scandurra, V., Stanfield, A. C., Zander, E., & Charman, T. (2018). Evaluating sex and age differences in ADI-R and ADOS scores in a large European multi-site sample of individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-018-3510-4

Uljarevic, M., Richdale, A. L., Evans, D. W., Cai, R. Y., & Leekam, S. R. (2017). Interrelationship between insistence on sameness, effortful control and anxiety in adolescents and young adults with autism spectrum disorder (ASD). *Molecular Autism, 8*(36). https://doi.org/10.1186/s13229-017-0158-4

van Wijngaarden-Cremers, P. J. M., van Eeten, E., Groen, W. B., van Deurzen, P. A., Oosterling, I. J., & van der Gaag, R. (2014). Gender and age differences in the core triad of impairments in autism spectrum disorders: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders, 44*, 627-635. https://doi.org/0.1007/s10803-013-1913-9

Vine Foggo, R. S., & Webster, A. A. (2017). Understanding the social experiences of adolescent females on the autism spectrum. *Research in Autism Spectrum Disorders, 35*, 74-84. https://doi.org/10.1016/j.rasd.2016.11.006

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2017). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*, 35-57. https://doi.org/10.3758/s13423-017-1343-3

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence- Third Edition*. [Measurement Instrument]. The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children- Fourth Edition*. [Measurement Instrument]. The Psychological Corporation.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence- Fourth Edition*. [Measurement Instrument]. The Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children- Fifth Edition*. [Measurement Instrument]. The Psychological Corporation.

Werling, D. M., & Geschwind, D. H. (2013). Understanding sex bias in autism spectrum disorder. *Proceedings of the Natlonal Academy of Sciences, 110*(13), 4868-4869. https://doi.org/10.1073/pnas.1301602110

Whitlock, A., Fulton, K., Lai, M. C., Pellicano, E., & Mandy, W. (2020). Recognition of girls on the autism spectrum by primary school educators: An experimental study. https://doi.org/10.1002/aur.2316

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wigham, S., Rogers, J., South, M., McConachie, H., & Freeston, M. (2015). The interplay between sensory processing abnormalities, intolerance of uncertainty, anxiety and

restricted and repetitive behaviours in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*(4), 943-954. https://doi.org/10.1007/s10803-014-2248-x

Wilke, C. O. (2016). *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=cowplot

Wilson, C. E., Murphy, C. M., McAlonan, G., Robertson, D. M., Spain, D., Hayward, H., Woodhouse, E., Deeley, P. Q., Gillan, N., Ohlsen, J. C., Zinkstok, J., Stoencheva, V., Faulkner, J., Yildiran, H., Bell, V., Hammond, N., Craig, M. C., & Murphy, D. G. M. (2016). Does sex influence the diagnostic evaluation of autism spectrum disorder in adults? *Autism, 20*(7), 808-819. https://doi.org/10.1177/1362361315611381

Wing, L. (1981). Sex ratios in early childhood autism and related conditions. *Psychiatry Research, 5*, 129-137.

Wood-Downie, H., Wong, B., Kovshoff, H., Mandy, W., Hull, L., & Hadwin, J. A. (2020). Sex/gender differences in camouflaging in children and adolescents with autism. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-020-04615-z

Worell, J., & Robinson, D. A. (2009). Issues in clinical assessment with women. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment*. Oxford University Press.

World Health Organisation. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. World Health Organisation.

Young, H., Oreve, M.-J., & Speranza, M. (2018). Clinical characteristics and problems diagnosing autism spectrum disorder in girls. *Archives de Pédiatrie, 25*(6), 399-403. https://doi.org/10.1016/j.arcped.2018.06.008

Young, R. L. (2007). *Autism Detection In Early Childhood (ADEC)*. ACER Press.

Young, R. L., & Rodi, M. L. (2014). Redefining Autism Spectrum Disorder using DSM-5: The implications of the proposed DSM-5 criteria for Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 44*, 758–765. https://doi.org/10.1007/s10803-013-1927-3

Appendices

Appendix A: Diagnostic Criteria for Autism Spectrum Disorder

Diagnostic and Statistical Manual of Mental Disorders, 5th Edition

(DSM-5; American Psychiatric Association, 2013)

A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays, and manifest by all three of the following:

1. Deficits in social-emotional reciprocity, ranging for example from abnormal social approach and failure of normal back and forth conversation; to reduced sharing of interests, emotions or affect; to failure to initiate or respond to social interactions.
2. Deficits in nonverbal communicative behaviours used for social interaction, ranging, for example from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expression and nonverbal communication.
3. Deficits in developing, maintaining and understanding relationships, ranging for example from difficulties adjusting behaviour to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers.

B. Restricted, repetitive patterns of behavior, interests, or activities as manifested by at least two of the following:

1. Stereotyped or repetitive motor movements, use of objects or speech (e.g., simple motor stereotypies, lining up toys of flipping objects, echolalia, idiosyncratic phrases).
2. Insistence on sameness, inflexible adherence to routines or ritualised patterns of verbal or nonverbal behaviour (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).
3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).
4. Hyper or hypo reactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching or objects, visual fascination with lights or movement).

C. Symptoms must be present in the early developmental period but may not become fully manifest until social demands exceed limited capabilities or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational or other important area of functioning.

E. These disturbances are not better explained by intellectual disability or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level.

Appendix B: Supplementary Age and IQ Descriptive Data (Study 1)

Table B.1

Sex/Gender Differences in Age and IQ of Children with GARS3 Item-Level Data Available

Variable	Males	Females	All
n	112	70	182
Age (years)			
M (SD)	8.72 (3.37)	8.87 (3.36)	8.78 (3.36)
Range	2.91-16.91	2.83-16.08	2.83-16.91
Full-scale IQ			
N available (% of children)	80 (71.43)	47 (67.14)	127 (69.78)
M (SD)	95.89 (16.93)	98.11 (13.57)	96.71 (15.75)
Range	40-134	70-131	40-134

Table B.2

Sex/Gender Differences in Age of Children with IQ Data Available

Variables	Instrument	Males	Females
	CARS2-ST	71 (39.7%)	27 (39.7%)
		5.93 (2.28)	6.70 (2.88)
n (% of total)	CARS2-HF	142 (66.4%)	69 (63.3%)
Age (years): *M (SD)*		9.05 (2.61)	9.76 (3.01)
	GARS-3	103 (63.2%)	64 (66.7%)
		8.92 (3.26)	9.32 (3.01)

Appendix C: Mathematical Formulations Underlying Statistical Models

The mathematical formulations underpinning the statistical models presented in this thesis are summarised here in order of appearance. This appendix was written with the statistical support of Associate Professor Nathan Weber.

Ordered Probit Model (Study 1)

Ordinal responses were modelled using an ordered probit model following the guidelines of Liddell and Kruschke (2018). Specifically, the ordinal ratings were modelled based on the categorisation of an underlying normally distributed variable with six thresholds used to determine the categorical response. The thresholds were modelled as invariant across levels of the predictors. To ensure model identifiability, the lowest and highest thresholds were anchored at 0.25 and 2.75 for the CARS2 data, and values of 1.25 and 3.75 were used for the GARS-3 data.

The mean (μ_g) of the latent variable (g) was estimated as the linear combination of the predictors: sex/gender, item, and their interaction, plus an intercept that was allowed to vary by client. The *SD* of the latent variable (σ_g) was estimated from the data and was given a uniform prior that was broad on the range of the rating scale (i.e., 4/1000 to 4 × 10).

$$g \sim N(\mu_g, \sigma_g)$$

$$\mu_g = b_0 + b_{client} + b_{sex} + b_{item} + b_{sex \times item}$$

$$\sigma_g \sim \text{uniform}(4/1000, 4 \times 10)$$

The intercept parameter (b_0) was given a normal prior with mean equal to the mid-point of the scale and standard deviation equal to the range of the rating scale. The other predictors, including the by-client deflections, were modelled as a vector of deflection parameters (one for each level of the predictor, or combination of predictors for the interaction). For each predictor, these were modelled as drawn from a normal distribution with a mean of 0 and *SD* estimated from the data, given a gamma distributed prior that was

based on the observed ratings. Thus, the deflection parameters (generically denoted as b_{factor}) were modelled as (note that normal distributions are parameterised a mean and standard deviation, not precision or variance, and gamma distribution as mode and *SD*):

$$b_0 \sim N(1.5, 4)$$

$$b_{factor} \sim N(0, \sigma_{factor})$$

$$\sigma_{factor} \sim gamma(sd_y/2, sd_y \times 2)$$

Finally, the free threshold parameters were given normal priors with mean equidistant between neighbouring ratings (e.g., for the threshold between 0.5 and 1, the prior was given mean of 0.75) and *SD* of 1 (i.e., twice the range of each rating category on the response scale). The probability of any given rating was then estimated using the cumulative normal distribution as the area under the normal curve between the relevant thresholds. The lower threshold for the first category was negative infinity and the higher threshold for the final category positive infinity.

_t_-tests (Studies 1, 2, and 3)

To conduct *t*-tests, the data was modelled from each group as *t*-distributed, with mean, scale, and normality parameter all estimated from the data. The mean was given a normal prior separate for each group, with mean equal to the observed mean and standard deviation vague on the observed scale (i.e., 100 times the observed *SD*). The scale parameter was given a uniform prior, also vague on the observed scale (observed *SD* divided by and multiplied by 1000 for the lower and upper bounds, respectively). Finally, the normality parameter (ν) was constrained to be greater than or equal to 1. This was achieved by modelling ν – 1 and giving this value an exponential prior with mean 29.

ANOVAs (Studies 2b and 3)

For ANOVAs with binary factors, a simple regression model was used. Specifically, dependent variable scores (*y*) were modelled as normal with mean estimated via linear link

function from an intercept parameter and parameters for each individual predictor, including all possible interactions (i.e., cross products). The standard deviation was estimated from the data and given a uniform prior that was vague on the standardised scale (i.e., 1/1000 to 1000). The predictor variables were always coded as 0.5 and -0.5 and the outcome variable standardised (to mean = 0 and *SD* = 1). The standard deviation of these normally distributed responses was also estimated from the data and was given a half-Cauchy prior with scale = 10 (vague on the scale of the standardised dependent variable). Parameters were allowed to vary by all relevant random effects (client in repeated measures designs, and also item/question when the design included multiple measurement items) by including a deflection parameter. These deflections were constrained to sum to zero (so the coefficient represents the effect averaged across all levels of the random effects) and were given normal prior with mean 0 and standard deviation estimated from the data. The standard deviations were given a half-Cauchy hyper prior with location 0 and scale 5.

Logistic Regressions (Studies 2 and 3)

Binary outcomes measures were analysed with a hierarchical Bayesian analogue of logistic regression. Specifically, the probability of a 1 (the target level of the outcome measure) was Bernoulli distributed with probability estimated, via the logit link function, as a linear combination of predictors. All models included an intercept and a coefficient for each predictor and, where relevant, all interaction terms. Binary predictors were coded as -0.5 and 0.5. The intercept was given a normal prior with mean 0 and standard deviation 10. All other coefficients were given normal priors with mean 0 and standard deviation 5. Where relevant, coefficients were allowed to vary by random effects (client/participant and/or question/item) by including a deflection parameter for all relevant coefficients. The deflection parameters were given a normal prior with mean 0 and standard deviation estimated from the data with a half-Cauchy hyperprior with location 0 and scale 5.

Nominal outcome variables with three levels were modelled using conditional logistic regression. Specifically, two logistic regression models were simultaneously fit to the data. The first predicting the probability of a level 1 versus levels 2 or 3) response. The second, predicting the conditional probability of a level 2 (versus 3) response, given a non-level 1 response. These two probabilities were then used (in each sample of the posterior) to estimate the probability of a response at each of the three levels. Each of the logistic regressions was constructed in the same manner as for binary outcome variables.

Appendix D: Supplementary Tables for Item-Level Analysis (Study 1)

Table D.1

Sex/Gender Differences in the Estimated Mean of the Latent Variable and Probability of Presenting with Impairment in CARS2-ST Items

CARS2-ST Item and Description (abbreviated)		Sex/Gender Difference in Estimated Mean of Latent Variable			Sex/Gender Difference in Probability Score ≥ 2		
		Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)
1. Relating to people	Response to communication, initiation of interaction, aloofness/ awareness of others	0.09 [0.01, 0.18]	.78	.20	0.03 [0.00, 0.06]	.66	.33
2. Imitation	Reliability, spontaneity, and immediacy of imitation, ability to imitate others	**0.19 [0.09, 0.28]**	**.98**	**.02**	**0.15 [0.08, 0.23]**	**.99**	**.01**
3. Emotional response	Appropriateness of emotion to situation, type/ degree of emotional response	0.04 [-0.04, 0.13]	.44	.45	0.01 [-0.01, 0.02]	.17	.82
4. Body use	Motor peculiarities and movement stereotypies, clumsiness, coordination	0.13 [0.04, 0.21]	.85	.14	0.05 [0.01, 0.08]	.84	.16
5. Object use	Degree of interest in objects, focus on parts of objects, repetitive or inappropriate use	**0.14 [0.05, 0.23]**	**.95**	.05	0.05 [0.02, 0.08]	.90	.10
6. Adaptation to change	Response to changes in routine, transitioning	-0.01 [-0.10, 0.08]	-.33	.50	0.00 [-0.01, 0.01]	.01	.99
7. Visual response	Abnormality in eye contact, visual stereotypies, visual sensory behaviour	**0.21 [0.12, 0.30]**	**1.00**	**.00**	**0.10 [0.05, 0.14]**	**1.00**	**.00**
8. Listening response	Auditory hyper/hypo-sensitivity	0.11 [0.02, 0.19]	.83	.16	0.04 [0.00, 0.07]	.77	.23
9. Taste, smell, and touch response and use	Response to sensory stimulation, use of these sensory modalities	0.02 [-0.07, 0.11]	.38	.48	0.01 [-0.02, 0.03]	.24	.71

CARS2-ST Item and Description (abbreviated)		Sex/Gender Difference in Estimated Mean of Latent Variable			Sex/Gender Difference in Probability Score ≥ 2		
		Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)
10. Fear or nervousness	Degree and context of anxious/nervous response	-0.11 [-0.22, -0.02]	-.81	.18	-0.04 [-0.07, -0.01]	-.77	.22
11. Verbal communication	Unusual speech mannerisms, preoccupation with certain topics, repetitive speech	**0.18 [0.09, 0.27]**	**.98**	**.02**	**0.06 [0.03, 0.09]**	**.96**	**.04**
12. Nonverbal communication	Impairment in expression, interpretation of nonverbal communication	**0.14 [0.05, 0.23]**	**.91**	**.08**	**0.07 [0.02, 0.11]**	**.92**	**.07**
13. Activity level	Hyper/hypo-activity	0.12 [0.03, 0.21]	.88	.12	**0.10 [0.03, 0.16]**	**.93**	**.05**
14. Level and consistency of intellectual response	General level of intellectual functioning, consistency in cognitive abilities	0.05 [-0.04, 0.15]	.61	.33	0.05 [-0.02, 0.12]	.72	.18
15. General impressions	Clinical impression of ASD severity	**0.13 [0.05, 0.22]**	**.92**	**.08**	0.05 [0.02, 0.08]	.89	.11

Note. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. Negative values indicate greater abnormality amongst females. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater impairment in males; negative probabilities indicate greater impairment in females). $P_{(within\ ROPE)}$ = probability that difference between sexes/genders was within the negligible range.

Table D.2

Sex/Gender Differences in the Estimated Mean of the Latent Variable and Probability of Presenting with Impairment in CARS2-HF Items

CARS2-HF Item and Description (abbreviated)		Sex/Gender Difference in Estimated Mean of Latent Variable			Sex/Gender Difference in Probability Score ≥ 2		
		Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)
1. Social-emotional Understanding	Understanding of non-verbal cues, perspectives of others	-0.01 [-0.07, 0.05]	.16	.70	0.00 [-0.87, 0.93]	.01	.99
2. Emotional expression and regulation	Appropriateness of type, degree of emotion, emotional reg., understanding	0.02 [-0.04, 0.08]	.22	.69	0.00 [-0.44, 0.66]	.00	1.00
3. Relating to people	Initiation of interaction, reciprocity of interactions	-0.03 [-0.09, 0.03]	-.30	.64	0.00 [-1.74, 0.80]	-.07	.92
4. Body use	Motor peculiarities, movement stereotypies, clumsiness, fine/gross motor skills	0.08 [0.01, 0.15]	.64	.38	0.03 [-0.41, 6.48]	.66	.33
5. Object use in play	Interest in toys or objects, repetitive or inappropriate use, imagination/ spontaneity in play	0.08 [0.00, 0.15]	.66	.34	0.04 [-0.44, 7.97]	.72	.27
6. Adaptation to change/restricted interests	Special and limited interests, rituals, routines, and ability to cope with change and transitions	-0.05 [-0.11, 0.01]	-.60	.40	0.00 [-0.95, 0.06]	.00	1.00
7. Visual response	Abnormality in eye contact, gaze switching, visual stereotypies, visual sensory behaviour	-0.01 [-0.07, 0.05]	-.26	.66	-0.01 [-3.72, 2.00]	-.28	.62
8. Listening response	Auditory hyper/hypo-sensitivity, response to name	-0.01 [-0.07, 0.04]	-.31	.63	-0.01 [-3.12, 1.34]	-.26	.70

CARS2-HF Item and Description (abbreviated)		Sex/Gender Difference in Estimated Mean of Latent Variable			Sex/Gender Difference in Probability Score ≥ 2		
		Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)
9. Taste, smell, and touch response and use	Response to sensory stimulation, use of these sensory modalities	-0.02 [-0.09, 0.03]	-.35	.60	-0.01 [-1.92, 0.68]	-.10	.89
10. Fear or anxiety	Extent of unusual fear or anxiety relative to context	-0.07 [-0.13, 0.00]	-.47	.51	-0.01 [-3.39, 0.62]	-.35	.64
11. Verbal communication	Verbal oddities, conversation reciprocity	0.02 [-0.04, 0.08]	.30	.64	0.01 [-0.98, 2.06]	.13	.86
12. Non-verbal communication	Use of facial expression and gestures, response to non-verbal behaviour, joint attention	-0.00 [-0.06, 0.06]	-.16	.70	0.00 [-1.65, 1.64]	.07	.87
13. Thinking and cognitive integration	Attention to detail, weak central coherence	0.00 [-0.06, 0.06]	-.18	.70	0.00 [-5.15, 4.32]	-.32	.43
14. Level and consistency of intellectual response	Overall intellectual functioning, consistency in cognitive abilities	0.03 [-0.03, 0.09]	.41	.56	0.03 [-2.17, 8.21]	.59	.32
15. General impressions	Clinical impression of ASD severity	-0.00 [-0.06, 0.06]	-.19	.69	0.00 [-2.19, 1.67]	-.12	.81

Note. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. Negative values indicate greater abnormality amongst females. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater impairment in males; negative probabilities indicate greater impairment in females). $P_{(within\ ROPE)}$ = probability that difference between sexes/genders was within the negligible range.

Table D.3

Sex/Gender Differences in Effect Size, Estimated Mean of the Latent Variable and Probability of Presenting with Impairment in all GARS3 Items

		Sex/Gender Difference in Estimated Mean of Latent Variable			Sex/Gender Difference in Probability Score ≥ 2		
GARS-3 Item	*d* [$HDI_{80\%}$]	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)
Restricted/Repetitive Behaviours (RB)							
1. If left alone, the majority of the individual's time will be spent in repetitive or stereotypical behaviours	0.10 [-0.03, 0.24]	0.12 [-0.04, 0.28]	.49	.47	0.04 [-0.01, 0.08]	.66	.27
2. Is preoccupied with specific stimuli that are abnormal in intensity	0.15 [0.02, 0.23]	0.18 [0.02, 0.35]	.69	.30	0.06 [0.01, 0.11]	.83	.14
3. Stares at hands, objects or items in the environment for at least 5 seconds	- 0.02 [- 0.17, 0.13]	-0.03 [-0.21, -0.15]	.13	.61	-0.01 [-0.07, 0.05]	-.40	.34
4. Flicks fingers rapidly in front of eyes for periods of 5 seconds or more	0.11 [-0.04, .26]	0.14 [-0.04, 0.31]	.55	.41	0.01 [0.00, 0.03]	.22	.77
5. Makes rapid lunging, darting movements when moving from place to place	0.17 [0.04, 0.31]	0.21 [0.05, 0.38]	.76	.23	0.07 [0.01. 0.12]	.88	.10
6. Flaps hands or fingers in front of face or at sides	0.17 [0.03, 0.31]	0.20 [0.04, 0.38]	.73	.26	0.05 [0.01, 0.09]	.82	.16
7. Makes high-pitched sounds (e.g., eee-eee-eee) or other vocalisations for self-stimulation	**0.28 [0.13, 0.43]**	**0.34 [0.15, 0.52]**	**.95**	**.05**	**0.11 [0.05, 0.17]**	**.98**	**.01**
8. Uses toys or objects inappropriately (e.g., spins cars, takes action toys apart)	**0.26 [0.11, 0.41]**	**0.32 [0.14, 0.50]**	**.94**	**.06**	**0.10 [0.04, 0.15]**	**.98**	**.02**
9. Does certain things repetitively, ritualistically	0.00 [-0.14, 0.15]	0.00 [-0.17, 0.18]	.18	.62	0.00 [-0.05, 0.05]	.32	.36

		Sex/Gender Difference in Estimated Mean of Latent Variable			Sex/Gender Difference in Probability Score ≥ 2		
GARS-3 Item	*d* [$HDI_{80\%}$]	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)	Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)
10. Engages in stereotyped behaviours when playing with toys or objects	0.10 [-0.03, 0.23]	0.13 [-0.04, 0.28]	.51	.46	0.04 [-0.01, 0.09]	.69	.24
11. Repeats unintelligible sounds (babbles) over and over	0.17 [0.04, 0.32]	0.21 [0.05, 0.39]	.76	.23	0.06 [0.01, 0.11]	.87	.11
12. Shows unusual interest in sensory aspects of play materials, body parts or objects	0.00 [-0.15, 0.14]	0.00 [-0.17, 0.18]	.18	.62	0.00 [-0.05, 0.06]	.32	.35
13. Displays ritualistic or compulsive behaviours	-0.02 [-0.16, 0.13]	-0.02 [-0.20, 0.16]	.14	.62	-0.01 [-0.06, 0.05]	-.37	.36
Social Interaction (SI)							
14. Does not initiate conversation with peers or others	0.10 [-0.04, 0.23]	0.12 [-0.05, 0.28]	.49	.47	0.03 [-0.01, 0.08]	.65	.27
15. Pays little or no attention to what peers are doing	0.14 [0.01, 0.28]	0.18 [0.02, 0.34]	.67	.32	0.06 [0.01, 0.11]	.82	.15
16. Fails to imitate other people in games or learning activities	0.16 [0.03, 0.30]	0.19 [0.03, 0.35]	.72	.28	0.06 [0.01, 0.12]	.86	.12
17. Does not follow others' gestures (cues) to look at something (e.g., when other person nods head, points, or uses other body language cues)	**0.24 [0.10, 0.38]**	**0.30 [0.13, 0.47]**	**.92**	**.08**	**0.09 [0.04, 0.14]**	**.97**	**.03**
18. Seems indifferent to other person's attention (doesn't try to get, maintain or direct the other person's attention)	0.18 [0.05, 0.32]	0.22 [0.06, 0.39]	.79	.20	0.07 [0.02, 0.13]	.90	.09
19. Shows minimal expressed pleasure when interacting with others	0.13 [0.00, 0.26]	0.15 [-0.01, 0.32]	.60	.38	0.05 [-0.01, 0.10]	.76	.19

GARS-3 Item	*d* [HDI$_{80\%}$]	Sex/Gender Difference in Estimated Mean of Latent Variable: Difference [HDI$_{80\%}$]	P (meaningful)	P (within ROPE)	Sex/Gender Difference in Probability Score ≥ 2: Difference [HDI$_{80\%}$]	P (meaningful)	P (within ROPE)
20. Displays little or no excitement in showing toys or objects to others	0.15 [0.01, 0.28]	0.18 [0.02, 0.34]	.68	.31	0.05 [0.00, 0.09]	.81	.16
21. Seems uninterested in pointing out things in the environment to others	0.14 [0.00, 0.27]	0.17 [0.01, 0.33]	.64	.34	0.05 [0.00, 0.11]	.80	.17
22. Seems unwilling or reluctant to get others to interact with him or her	0.17 [0.03, 0.30]	0.21 [0.04, 0.36]	.75	.24	0.07 [0.01, 0.12]	.88	.11
23. Shows minimal or no response when others attempt to interact with him or her	0.10 [-0.04, 0.23]	0.13 [-0.04, 0.28]	.51	.46	0.04 [-0.02, 0.09]	.69	.23
24. Displays little or no reciprocal communication (e.g., doesn't voluntarily say 'bye-bye' in response to another person saying 'bye-bye' to him or her)	0.20 [0.06, 0.34]	0.24 [0.08, 0.41]	.83	.17	**0.08 [0.03, 0.13]**	**.92**	**.07**
25. Doesn't try to make friends with people	0.21 [0.07, 0.34]	0.25 [0.08, 0.42]	.85	.15	**0.08 [0.02, 0.13]**	**.94**	**.06**
26. Fails to engage in creative, imaginative play	**0.25 [0.11, 0.39]**	**0.30 [0.13, 0.47]**	**.92**	**.08**	**0.10 [0.04, 0.15]**	**.97**	**.03**
27. Shows little or no interest in other people	0.19 [0.05, 0.32]	0.23 [0.07, 0.39]	.82	.18	**0.08 [0.02, 0.13]**	**.92**	**.07**
Social Communication (SC)							
28. Responds inappropriately to humorous stimuli (e.g., doesn't laugh at jokes, cartoons, funny stories)	0.02 [-0.12, 0.16]	0.02 [-0.15, 0.20]	.22	.63	0.01 [-0.05, 0.06]	.39	.34
29. Has difficulty understanding jokes	0.10 [-0.04, 0.23]	0.12 [-0.05, 0.28]	.49	.48	0.03 [-0.01, 0.08]	.65	.28
30. Has difficulty understanding slang expressions	0.09 [-0.04, 0.23]	0.11 [-0.05, 0.28]	.47	.49	0.03 [-0.02, 0.07]	.62	.30

GARS-3 Item	*d* [HDI$_{80\%}$]	Sex/Gender Difference in Estimated Mean of Latent Variable: Difference [HDI$_{80\%}$]	P (meaningful)	P (within ROPE)	Sex/Gender Difference in Probability Score ≥ 2: Difference [HDI$_{80\%}$]	P (meaningful)	P (within ROPE)
31. Has difficulty identifying when someone is teasing	0.16 [0.03, 0.30]	0.20 [0.03, 0.36]	.73	.27	0.05 [0.01, 0.09]	.83	.15
32. Has difficulty understanding when he or she is being ridiculed	0.09 [-0.05, 0.23]	0.10 [-0.06, 0.27]	.44	.51	0.03 [-0.02, 0.07]	.59	.32
33. Has difficulty understanding what causes people to dislike him or her	0.08 [-0.06, 0.23]	0.10 [-0.08, 0.27]	.43	.51	0.01 [-0.01, 0.08]	.26	.71
34. Fails to predict probable consequences in social events	0.11 [-0.03, 0.25]	0.14 [-0.04, 0.30]	.54	.43	0.02 [-0.01, 0.09]	.49	.49
35. Doesn't seem to understand that people have thoughts and feelings different from his or hers	0.01 [-0.14, 0.16]	0.01 [-0.17, 0.19]	.20	.61	0.00 [-0.03, 0.03]	.20	.63
36. Doesn't seem to understand that the other person doesn't know something	0.12 [-0.02, 0.26]	0.14 [-0.02, 0.32]	.57	.40	0.03 [0.00, 0.06]	.63	.34
Emotional Responses (ER)							
37. Needs an excessive amount of reassurance if things are changed or go wrong	0.09 [-0.06, 0.23]	0.10 [-0.08, 0.28]	.44	.49	0.01 [-0.01, 0.04]	.34	.62
38. Becomes frustrated quickly when he or she cannot do something	0.11 [-0.03, 0.26]	0.14 [-0.05, 0.31]	.54	.42	0.01 [0.00, 0.02]	.11	.89
39. Temper tantrums when frustrated	0.04 [-0.12, 0.19]	0.04 [-0.14, 0.23]	.29	.57	0.00 [-0.01, 0.02]	.06	.91
40. Becomes upset when routines are changed	0.09 [-0.06, 0.22]	0.11 [-0.07, 0.27]	.45	.50	0.02 [-0.01, 0.05]	.50	.44

GARS-3 Item	*d* [HDI$_{80\%}$]	Sex/Gender Difference in Estimated Mean of Latent Variable: Difference [HDI$_{80\%}$]	P (meaningful)	P (within ROPE)	Sex/Gender Difference in Probability Score ≥ 2: Difference [HDI$_{80\%}$]	P (meaningful)	P (within ROPE)
41. Responds negatively when given commands, requests or directions	0.06 [-0.08, 0.20]	0.07 [-0.10, 0.24]	.35	.57	0.01 [-0.02, 0.05]	.43	.46
42. Has extreme reactions (e.g., cries, screams, tantrums) in response to loud, unexpected noise	0.09 [-0.05, 0.22]	0.11 [-0.06, 0.27]	.45	.51	0.03 [-0.01, 0.09]	.64	.26
43. Temper tantrums when doesn't get his or her way	0.01 [-0.14, 0.16]	0.01 [-0.17, 0.19]	.21	.60	0.00 [-0.03, 0.03]	.22	.60
44. Temper tantrums when told to stop doing something he or she enjoys doing	0.10 [-0.04, 0.24]	0.12 [-0.05, 0.29]	.50	.46	0.02 [-0.01, 0.05]	.54	.42
Cognitive Style (CS)							
45. Uses exceptionally precise speech	0.13 [-0.01, 0.26]	0.15 [-0.01, 0.32]	.60	.38	0.05 [0.00, 0.10]	.77	.18
46. Attaches very concrete meaning to words	0.15 [0.02, 0.29]	0.18 [0.02, 0.34]	.68	.31	0.05 [0.00, 0.10]	.81	.16
47. Talks about a single subject excessively	0.15 [0.01, 0.28]	0.18 [0.01, 0.34]	.67	.32	0.05 [0.00, 0.09]	.80	.18
48. Displays superior knowledge or skill in specific subjects	**0.35 [0.18, 0.50]**	**0.42 [0.21, 0.61]**	**.99**	**.01**	**0.12 [0.06, 0.18]**	**1.00**	**.00**
49. Displays excellent memory	0.15 [0.01, 0.28]	0.18 [0.01, 0.34]	.68	.30	0.04 [0.00, 0.09]	.79	.19
50. Shows an intense, obsessive interest in specific intellectual subjects	**0.32 [0.16, 0.48]**	**0.39 [0.20, 0.58]**	**.98**	**.02**	**0.12 [0.06, 0.18]**	**.99**	**.01**
51. Makes naïve remarks (unaware of reaction produced in others)	0.02 [-0.13, 0.16]	0.02 [-0.15, 0.19]	.21	.62	0.01 [-0.04, 0.05]	.34	.40

GARS-3 Item	*d* [$HDI_{80\%}$]	Sex/Gender Difference in Estimated Mean of Latent Variable: Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)	Sex/Gender Difference in Probability Score ≥ 2: Difference [$HDI_{80\%}$]	P (meaningful)	P (within ROPE)
Maladaptive Speech (MS)							
52. Repeats (echoes) words or phrases verbally or with signs	0.09 [-0.04, 0.23]	0.11 [-0.06, 0.27]	.47	.48	0.04 [-0.02, 0.09]	.65	.25
53. Repeats words out of context (repeats words or phrases heard at an earlier time)	0.15 [0.01, 0.28]	0.18 [0.01, 0.34]	.67	.32	0.06 [0.00, 0.11]	.82	.15
54. Speaks (or signs) with flat tone, affect	0.15 [0.02, 0.29]	0.19 [0.03, 0.37]	.70	.29	0.06 [0.01, 0.11]	.83	.15
55. Uses 'yes' or 'no' inappropriately. Says 'yes' when asked if he or she wants an aversive stimulus or says 'no' when asked if he or she wants a favorite toy or treat	0.16 [0.01, 0.29]	0.19 [0.02, 0.36]	.70	.29	0.04 [0.00, 0.07]	.75	.24
56. Uses 'he' or 'she' instead of 'I' when referring to self	0.16 [0.01, 0.30]	0.19 [0.02, 0.38]	.70	.29	0.02 [0.00, 0.04]	.56	.43
57. Speech is abnormal in tone, volume or rate	0.17 [0.04, 0.31]	0.21 [0.05, 0.38]	.77	.23	0.07 [0.01, 0.12]	.88	.10
58. Utters idiosyncratic words or phrases that have no meaning to others	0.17 [0.03, 0.31]	0.20 [0.04, 0.38]	.74	.25	0.05 [0.01, 0.10]	.84	.14

Note. Negative values indicate greater abnormality amongst females. Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the negligible range and in the observed direction (positive probabilities indicate greater impairment in males; negative probabilities indicate greater impairment in females). $P_{(within\ ROPE)}$ = probability that difference between sexes/genders was within the negligible range. *d* reflects the effect size of the male-female difference in the estimated latent means.

Table D.4

Comparisons of Cohen's d Statistics: Study 1 and Kumazaki et al. (2015)

	Study 1		Kumazaki et al. (2015)	
	d [$HDI_{80\%}$]	$P_{(meaningful)}$	*d*	Higher group
1. Relating to people	0.21 [0.03, 0.38]	.78	0.20	Female
2. Imitation	**0.38 [0.19, 0.57]**	**.98**	0.38	Male
3. Emotional response	0.08 [-0.10, 0.26]	.44	0.03	Male
4. Body use	0.24 [0.07, 0.42]	.85	0.68†	Male
5. Object use	**0.33 [0.15, 0.51]**	**.95**	1.11*	Male
6. Adaptation to change	-0.03 [-0.22, 0.15]	-.33	0.44	Male
7. Visual response	**0.47 [0.28, 0.66]**	**1.00**	0.41	Female
8. Listening response	0.23 [0.05, 0.41]	.83	0.43	Female
9. Taste, smell, touch response and use	0.06 [-0.13, 0.23]	.38	0.86*	Female
10. Fear or nervousness	-0.24 [-0.45, -0.03]	-.81	0.75†	Female
11. Verbal communication	**0.39 [0.21, 0.57]**	**.98**	0.40	Male
12. Nonverbal communication	**0.29 [0.11, 0.47]**	**.91**	0.30	Male
13. Activity level	0.26 [0.08, 0.44]	.88	0.71†	Male
14. Level and consistency of intellectual response	0.14 [-0.05, 0.33]	.61	0.58	Female
15. General impressions	**0.30 [0.11, 0.47]**	**.92**	0.30	Female

Note. † $p < .05$, * $p < .01$.

Appendix E: Operationalisation of Behaviours (Study 2a and 2b)

Table E

Operationalisation of Behaviours Examined in Study 2a and 2b

Criterion: Source	Behavioural category	Operationalisation
Criterion A1	***Deficits in socio-emotional reciprocity***	**Concern = 2, some/partial concern = 1, no concern = 0 for each group of behaviours listed below**
Parent report	Social approach	Atypical initiation, response, appearing 'in own world, inappropriate timing of approach, approaching strangers
	Social norms	Lack of awareness of boundaries/personal space, social manners, socially inappropriate comments
	Reciprocal conversation	Difficulties with conversation reciprocity, e.g., turn taking, asking questions, interrupting
	Sharing interests	Difficulty sharing in a topic of conversation, joint attention
	Sharing emotion	Difficulty sharing in others' emotions, reciprocal smiling, shared enjoyment
	Conversation content	Unusual conversation features, e.g., excessive detail, monologing, redirecting topic
	Literal language	Literal interpretation of jokes/sarcasm
Diagnostic observation	As above	As above
Criterion A2	***Deficits in nonverbal communicative behaviours used for social interaction***	**Concern = 2, some/partial concern = 1, no concern = 0 for each group of behaviours listed below**
Parent report	Integration of verbal/nonverbal behaviour	Difficulties with consistency between nonverbal and verbal behaviour (e.g., smiling when sounding angry)
	Eye contact	Atypicality including absence, staring, inconsistency
	Use of nonverbal communication	Atypicality including exaggerated or absent descriptive/social/emphatic gestures
	Facial expression	Difficulty including exaggerated, absent, or ill-suited to context
	Nonverbal understanding	Difficulties in recognising/understanding nonverbal communication
	Response to nonverbal behaviour	Difficulty in spontaneity, regularity and appropriateness of response to non-vernal behaviour

Criterion: Source	Behavioural category	Operationalisation
	Emotional regulation	Difficulty regulating emotions, comfort seeking
Diagnostic observation	Integration of verbal/nonverbal behaviour	Difficulties with consistency between nonverbal and verbal behaviour (e.g., smiling when sounding angry)
	Eye contact	Atypicality including absence, staring, inconsistency
	Use of nonverbal communication	Atypicality including exaggerated or absent descriptive/social/emphatic gestures
	Facial expression	Difficulty including exaggerated, absent, or ill-suited to context
	Nonverbal understanding	Difficulties in recognising/understanding nonverbal communication
Criterion A3	***Deficits in developing and maintaining relationships appropriate to developmental level***	**Concern = 2, some/partial concern = 1, no concern = 0 for each group of behaviours listed below**
Parent report	Adjusting behaviour for situation	Difficulty modifying behaviour to environment or the person with whom the child is interacting, understanding of social hierarchies
	Imagination/spontaneity in play	Difficulty displaying novel or imaginative play themes, use of play scripts
	Submissive/dominating in play	Tendency to allow others to direct the child in play, or to impose their own play ideas upon others
	Possessive/difficulty losing	Possessiveness of objects, difficulty losing a game
	Friendship formation	Difficulty making friends, initiating a relationship
	Friendship maintenance	Difficulty maintaining friendships, resolving conflict
	Social motivation	Absent or excessive motivation for friendships or social interaction
	Consistent companions	Lack of consistent companions, superficial companionship (e.g., parallel play)
Diagnostic observation	Friendship understanding	Difficulty understanding what a friend means, how to be a friend, interests of friends etc.
	Inclusiveness of assessor in play	Degree to which the child excluded the assessor in play, and appropriateness of play behaviour (e.g., submissive or dominating

Criterion: Source	Behavioural category	Operationalisation
	Imagination, spontaneity in play	Difficulty displaying novel or imaginative play themes, use of play scripts. For older children, engagement in non-clinical discussion or structured play (e.g., board games)
Criterion B1	***Stereotyped or repetitive speech, motor movement or use of objects***	**Concern = 2, some/partial concern = 1, no concern = 0 for each group of behaviours listed below**
Parent report	Stereotypical movement	Overall level of concern regarding stereotypical movement: pervasiveness, variety, frequency, and functional disruption of behaviours
	Specific motor behaviours	Specify (i.e., list number corresponding to the behaviour): 1. Toe walking 2. Flapping 3. Spinning 4. Gross motor mannerism (e.g., unusual gait) 5. Rocking/swaying/jumping 6. Rigidity (physical) 7. Hand mannerisms (e.g., twinkling or posturing) 8. Mouth mannerisms (e.g., grimacing) 9. Self-injurious (e.g., head banging) 10. Repetitive body use (e.g., thumb sucking, picking at skin)
	Stereotypical speech/ language	Overall level of concern regarding stereotypical speech: pervasiveness, variety, frequency, and functional disruption of behaviours
	Speech/language	Specify (i.e., list number corresponding to the behaviour): 1. Echolalia 2. Third person referencing 3. Neologisms/idiosyncratic speech 4. Pronoun reversal 5. Repetitive speech 6. Accents 7. Unusual noises, self-induced noises 8. Talking to self 9. Odd prosody/tone/volume
	Stereotypical object use	Overall level of concern regarding stereotypical object use: pervasiveness, variety, frequency, and functional disruption of behaviours

Criterion: Source	Behavioural category	Operationalisation
	Object use	Specify (i.e., list number corresponding to the behaviour): 1. Lining up 2. Grouping 3. Spinning/flicking/pushing 4. Repetitive play/ object use 5. Deconstructing/attention to parts of objects
Diagnostic observation	Stereotypical movement	Overall level of concern regarding stereotypical movement: pervasiveness, variety, frequency, and functional disruption of behaviours
	Stereotypical speech	Overall level of concern regarding stereotypical speech: pervasiveness, variety, frequency, and functional disruption of behaviours
	Stereotypical object use	Overall level of concern regarding stereotypical object use: pervasiveness, variety, frequency, and functional disruption of behaviours
Criterion B2	***Excessive adherence to routines, ritualised patterns of verbal or nonverbal behaviour, or excessive resistance to change***	**Concern = 2, some/partial concern = 1, no concern = 0 for each group of behaviours listed below**
Parent report	Distress at change	Distress associated with novelty, change or unfamiliarity
	Routine adherence	Insistence on unusual or pedantic rituals that are not necessarily functional
	Task switching/transitioning	Difficulty switching tasks or transitioning between activities, need for completion
	Cognitive rigidity	Black and white thinking, rule adherence
Diagnostic observation	Difficulties with transitioning	Difficulty switching tasks or transitioning between activities, need for completion
	Cognitive rigidity	Black and white thinking, rule adherence
	Routine adherence	Development of unusual or pedantic rituals that are not necessarily functional

Criterion: Source	Behavioural category	Operationalisation
Criterion B3	***Highly restricted, fixated interests that are abnormal in intensity or focus.***	**Concern = 2, some/partial concern = 1, no concern = 0 for each group of behaviours listed below**
Parent report	Restricted interests	Specify (i.e., list number corresponding to the behaviour): 1. Specific program/character 2. Random objects (e.g., rocks, shells) 3. Vehicles (including toy vehicles) 4. Toys (not included above e.g., teddy bear) 5. Screens (e.g., video games) 6. Animals 7. Systems (e.g., numbers, routes, schedules) 8. Craft/art (e.g., drawing) 9. Sport/activity (e.g., sailing) 10. People (e.g., celebrities or someone known to the child)
Diagnostic observation	Restricted interests	Degree to which a restricted interest was apparent during assessment
Criterion B4	***Hyper or hypo-reactivity to sensory input or unusual interest in sensory aspects of the environment***	**Concern = 2, some/partial concern = 1, no concern = 0 for each group of behaviours listed below**
Parent report	Auditory	Seeking (e.g., loud music) Avoidant (e.g., covering ears, avoidance, fear)
	Tactile	Seeking (e.g., inappropriate touching, excessive tactile behaviours) Avoidant (e.g., refusal to touch certain textures, selection of clothing based on feel)
	Olfactory	Seeking (e.g., sniffing people/objects) Avoidant (e.g., commenting/gagging)
	Oral	Seeking (e.g., mouthing/chewing/licking objects/people) Avoidant (of certain foods, flavours)
	Visual	Seeking (e.g., looking at objects from unusual angles, fascination with particular visual experiences) Avoidant (e.g., sensitivity to sunlight)
Diagnostic observation	Sensory behaviour	Overall level of concern regarding sensory behaviours: pervasiveness, variety, frequency, and functional disruption of behaviours

Criterion: Source	Behavioural category	Operationalisation
Teacher report		
Other	Academic performance	0 = no concern 1 = behind in a specific class 2 = behind in all classes
Criterion A1	Social interest/approach	How interested is the child in his/her peers? 0 = no concern 1 = some concern: inconsistent interest or slight disinterest in peers 2 = significant concern: lack of interest in peers, excessive interest/obsessions with peers
	Conversation skills	Reciprocal conversation, interrupting, monologuing, conversation content 0 = no concern 1 = some concern: some difficulty or inconsistencies in some or all of the above 2 = significant concern: difficulties are pervasive and interfering
Criterion A2	Non-verbal interpretation	Interpretation of body language, emotions 0 = no concern 1 = some concern: some difficulty or inconsistencies in some or all of the above 2 = significant concern: difficulties are pervasive and interfering
	Use of nonverbal communication	Eye contact, facial expressions, gestures 0 = no concern 1 = some concern: some difficulty or inconsistencies in some or all of the above 2 = significant concern: difficulties are pervasive and interfering
Criterion A3	Friendship formation	To what extent is the child able to maintain friendships, negotiate conflicts, play appropriately? 0 = no concern 1 = some concern: inconsistencies in the above or mild difficulties 2 = significant concern: these difficulties are pervasive and interfering
	Friendship maintenance	To what extent is the child able to negotiate conflicts and maintain friendships? 0 = no concern 1 = some concern: inconsistencies in the above or mild difficulties 2 = significant concern: these difficulties are pervasive and interfering

Criterion: Source	Behavioural category	Operationalisation
Criterion B1	Stereotypical movement	0 = no concern 1 = some concern, some behaviours noted but these are not interfering 2 = significant concern, interfering behaviours noted
	Stereotypical speech	0 = no concern 1 = some concern, some behaviours noted but these are not interfering 2 = significant concern, interfering behaviours noted
	Stereotypical object use	0 = no concern 1 = some concern, some behaviours noted but these are not interfering 2 = significant concern, interfering behaviours noted
Criterion B2	Routines and rituals	Including motor and verbal 0 = no concern 1 = some concern: dislikes changes in routine and rituals present 2 = significant concern: cannot cope with changes in routine, rituals are marked and intrusive
	Difficulties with change	Including difficulties with transition 0 = no concern 1 = some concern, sometimes has difficulty, difficulty is mild and not overly interfering 2 = significant concern, interfering
Criterion B3	Restricted interests	Has a restricted interest/preoccupation been reported? 2 = yes 1 = somewhat (clear interest reported but unclear if restricted) 0 = no
Criterion B4	Sensory behaviours	0 = no concern 1 = some concern, sometimes has difficulty, difficulty is mild and not overly interfering 2 = significant concern, interfering

Appendix F: Frequency of Severity Ratings Across Domains (Study 2a and 2b) [66]

Figure F.1

Proportions of Severity Ratings for ASD Characteristic Across Assessments (Parent Report; Study 2a)

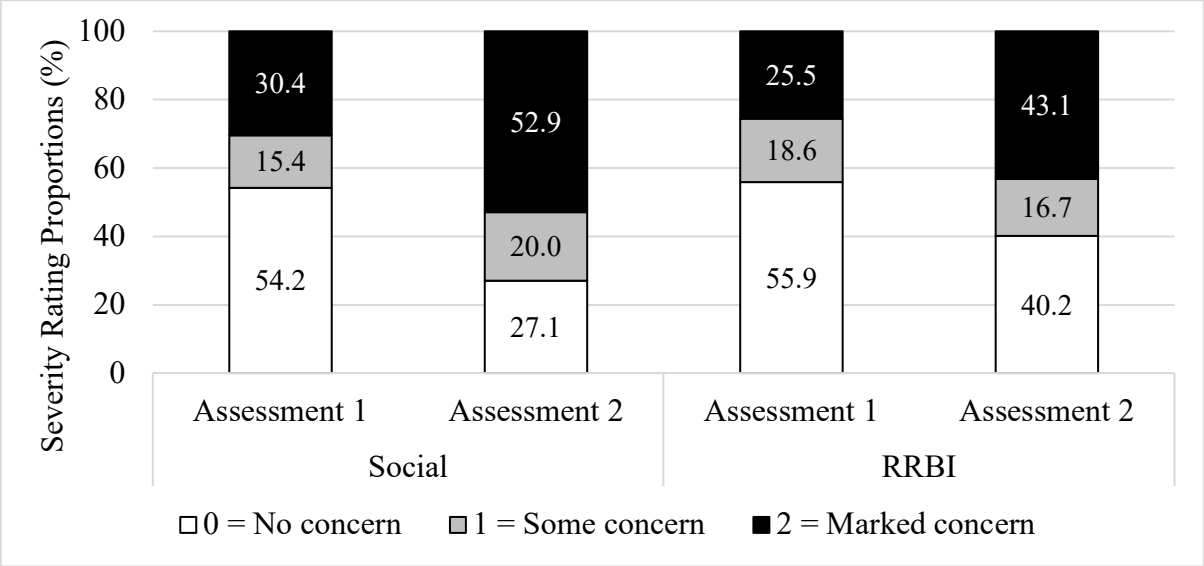
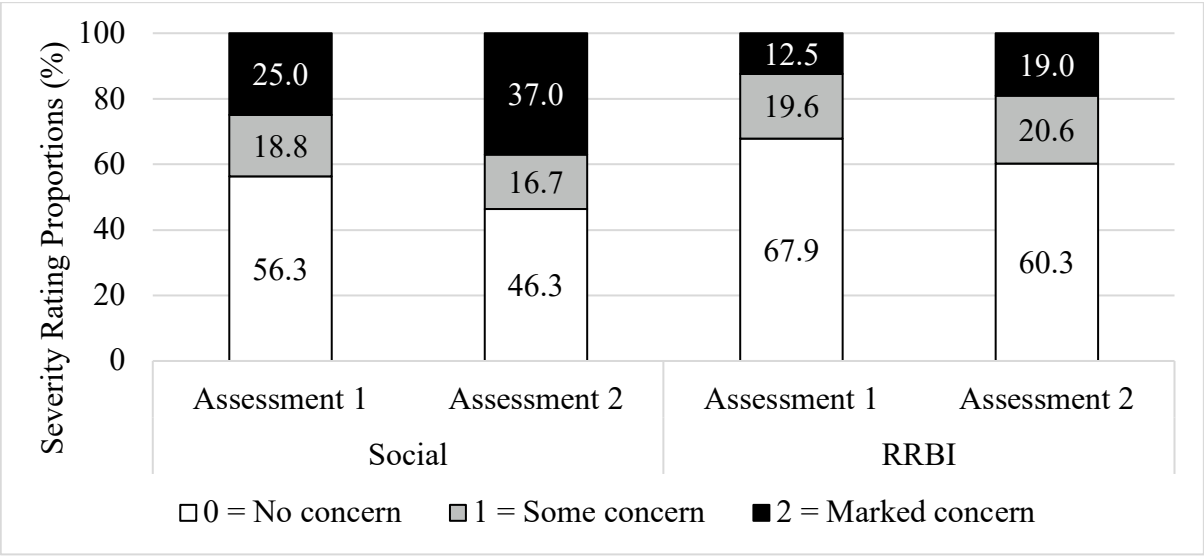


Figure F.2

Proportions of Severity Ratings for ASD Characteristic Across Assessments (Teacher Report; Study 2a)



[66] These figures were derived from the raw proportions data rather than the models. Therefore, HDIs (80%) are not included.

Figure F.3

Proportions of Severity Ratings for ASD Characteristic Across Assessments (Diagnostician Observation; Study 2a)

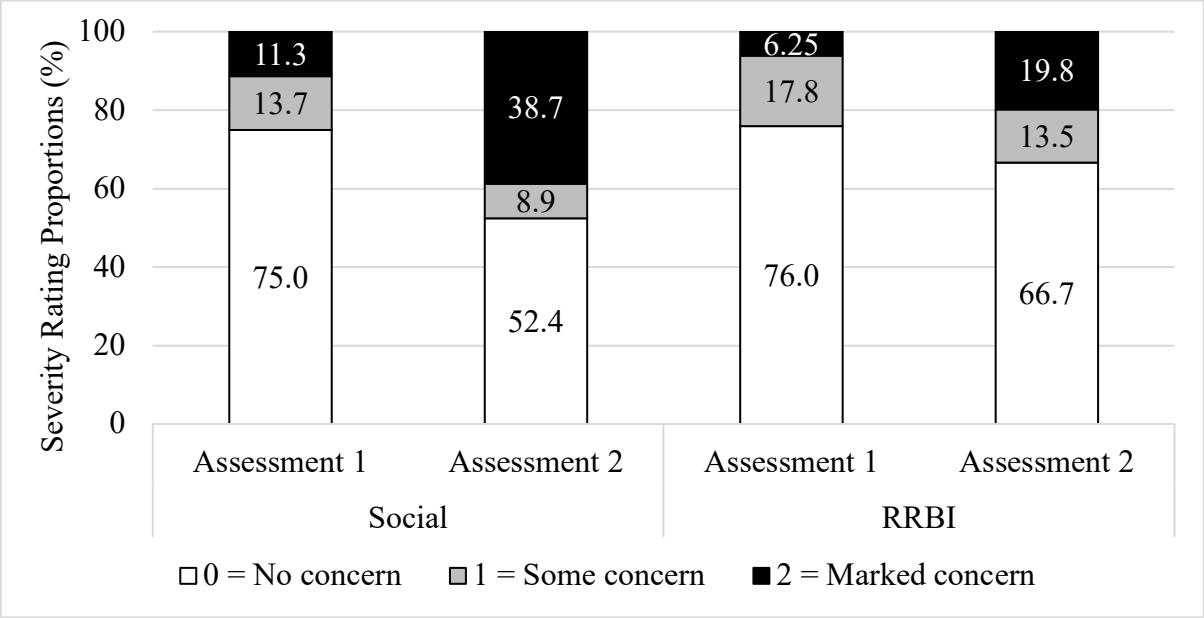


Figure F.4

Proportion of ASD Behaviours for Which Concern Was Reported by Parents (Study 2b)

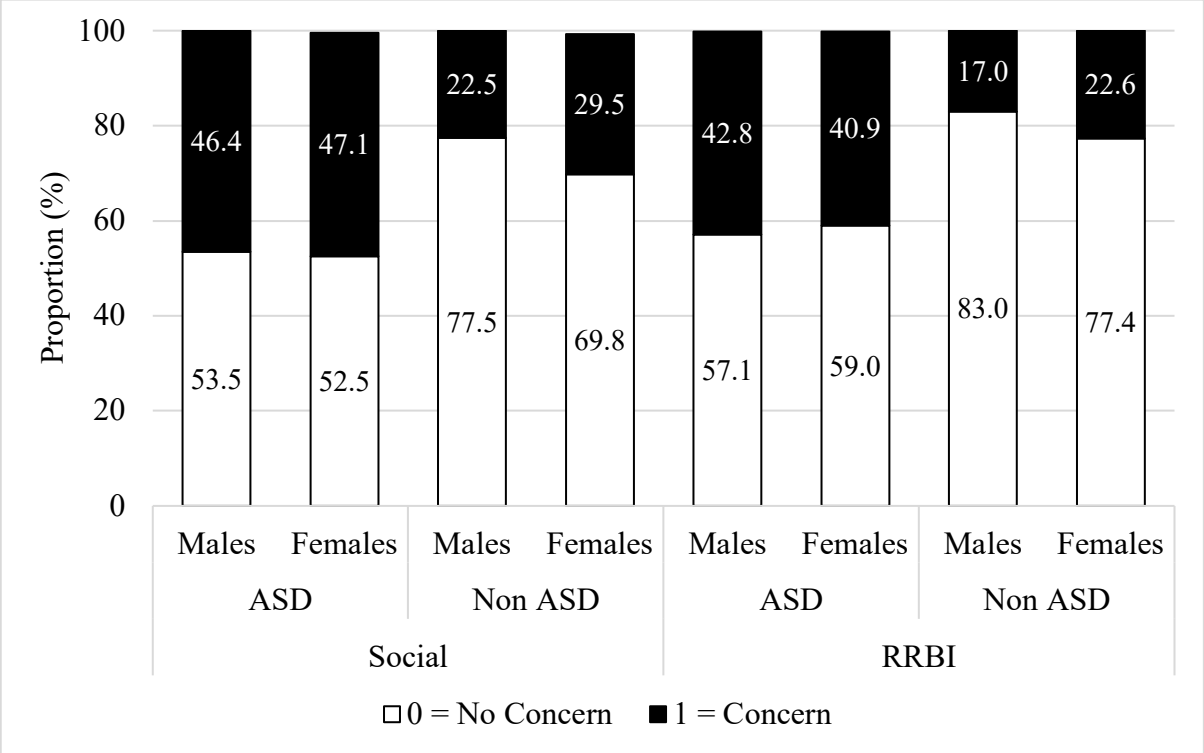


Figure F.5

Proportion of ASD Behaviours for Which Concern Was Observed by Diagnosticians (Study 2b)

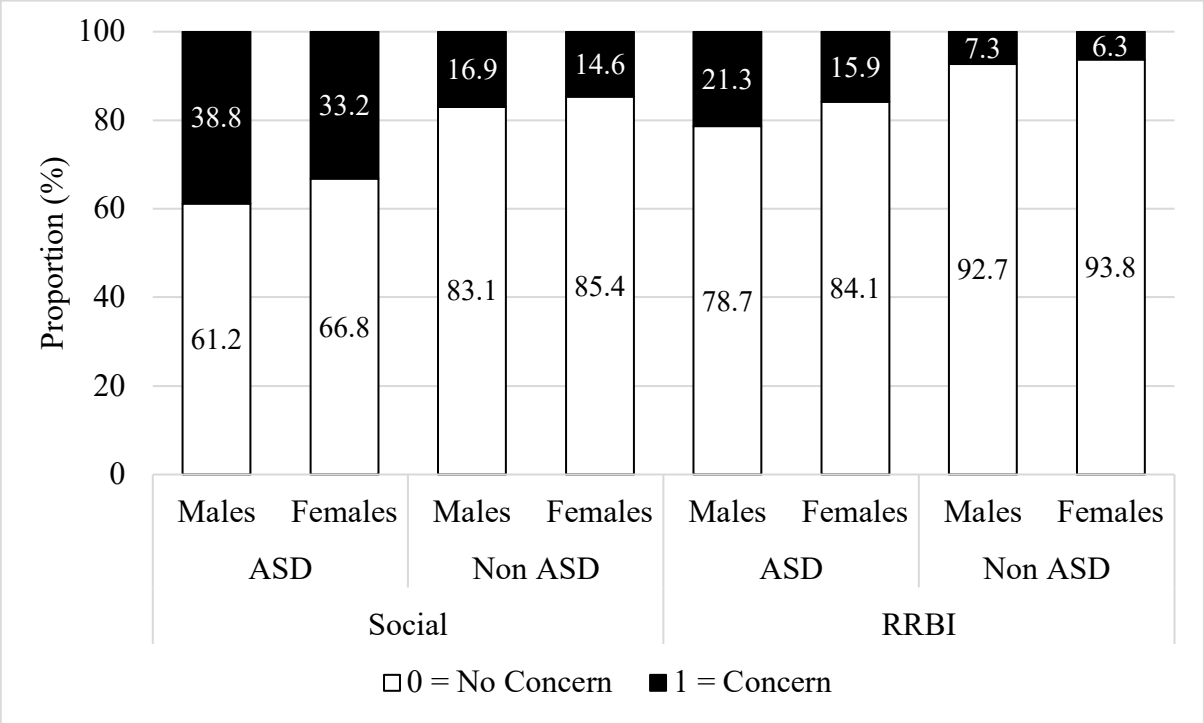
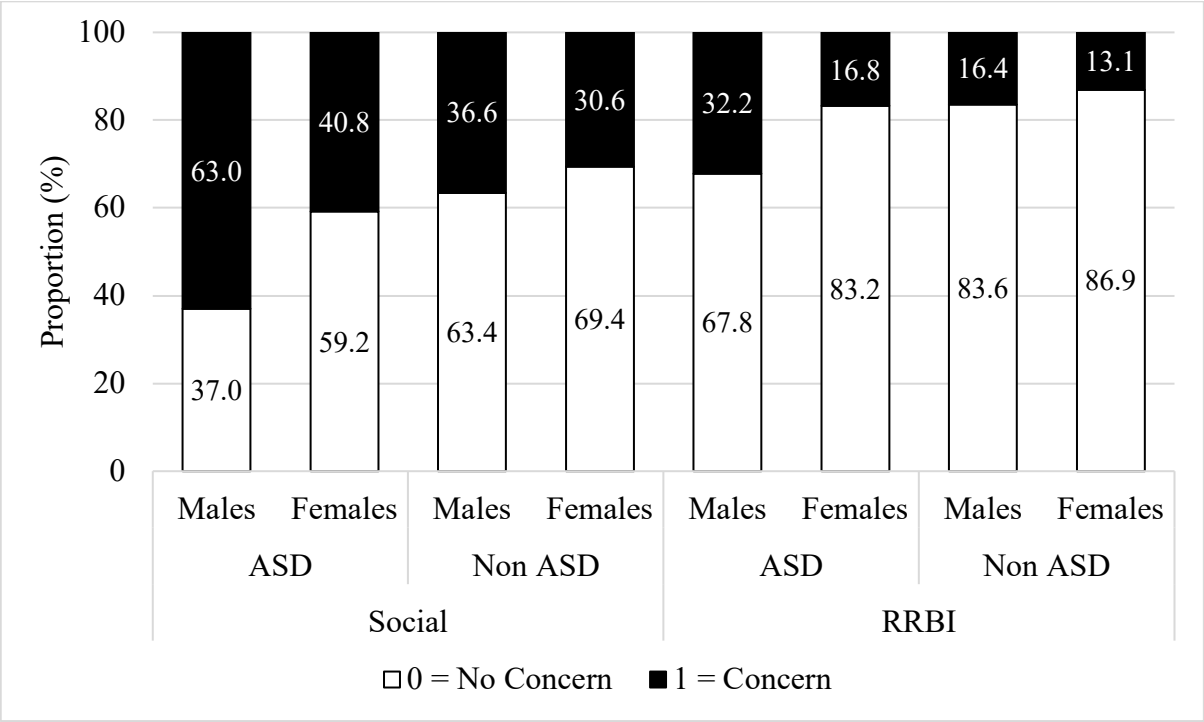


Figure F.6

Proportion of ASD Behaviours for Which Concern Was Reported by Teachers (Study 2b)



Appendix G: Supplementary Tables for Study 2b

Table G.1

Proportions of Participants for Whom Concern Was Raised for Each Behaviour

	Group: Proportion of total [$HDI_{80\%}$]			
	ASD		Non-ASD	
Behavioural Category	Male	Female	Male	Female
Social-emotional reciprocity				
Parent report				
Social approach	.77 [.67, .86]	.73 [.62, .82]	.39 [.26, .53]	.55 [.34, .76]
Social norms	.60 [.49, .71]	.61 [.49, .71]	.26 [.14, .39]	.40 [.20, .61]
Reciprocal conversation	.57 [.45, .68]	.67 [.56, .77]	.28 [.16, .41]	.24 [.08, .43]
Sharing interests	.19 [.05, .37]	.30 [.20, .40]	.06 [.01, .14]	.29 [.19, .39]
Sharing emotions	.27 [.17, .37]	.25 [.16, .35]	.10 [.03, .20]	.24 [.08, .43]
Content of conversation	.73 [.62, .82]	.69 [.59, .80]	.30 [.18, .44]	.34 [.15, .55]
Literal language	.18 [.10, .28]	.29 [.19, .40]	.19 [.09, .31]	.34 [.15, .55]
Diagnostician observations				
Social approach	.39 [.28, .50]	.38 [.28, .50]	.10 [.03, .19]	.13 [.02, .29]
Social norms	.42 [.30, .52]	.34 [.24, .46]	.21 [.10, .33]	.13 [.02, .29]
Reciprocal conversation	.77 [.67, .86]	.61 [.50, .72]	.30 [.18, .44]	.24 [.08, .43]
Sharing interests	.17 [.09, .26]	.11 [.05, .19]	.07 [.01, .15]	.00 [.00, .05]
Sharing emotions	.10 [.04, .17]	.11 [.05, .19]	.05 [.01, .12]	.03 [.00, .13]
Content of conversation	.55 [.44, .66]	.58 [.46, .69]	.27 [.15, .41]	.04 [.00, .15]
Literal language	.09 [.03, .17]	.04 [.01, .09]	.18 [.08, .30]	.12 [.01, .27]
Teacher report				
Academic achievement	.25 [.13, .38]	.37 [.22, .51]	.24 [.12, .39]	.39 [.12, .68]
Social approach	.61 [.48, .74]	.41 [.27, .55]	.45 [.30, .61]	.41 [.16, .68]
Reciprocal conversation	.68 [.54, .80]	.39 [.26, .53]	.35 [.21, .50]	.24 [.04, .48]
Nonverbal communication				
Parent report				
Integration of verbal/ nonverbal behaviour	.14 [.07, .22]	.30 [.19, .40]	.00 [.00, .02]	.03 [.00, .13]
Eye contact	.54 [.42, .65]	.62 [.51, .73]	.26 [.14, .39]	.40 [.19, .61]
Use of nonverbal communication	.30 [.20, .40]	.25 [.16, .35]	.10 [.03, .20]	.19 [.05, .37]
Facial expression	.37 [.26, .48]	.63 [.52, .74]	.10 [.03, .20]	.29 [.11, .49]
Nonverbal understanding	.63 [.52, .74]	.70 [.60, .80]	.28 [.16, .41]	.44 [.24, .66]
Response to nonverbal behaviour	.43 [.32, .55]	.46 [.35, .58]	.15 [.06, .26]	.19 [.05, .37]

	Group: Proportion of total [$HDI_{80\%}$]			
	ASD		Non-ASD	
Behavioural Category	Male	Female	Male	Female
Emotional regulation	.58 [.46, .69]	.57 [.45, .68]	.41 [.27, .55]	.34 [.16, .55]
Diagnostician observations				
Eye contact	.47 [.36, .59]	.43 [.32, .55]	.17 [.07, .28]	.19 [.05, .37]
Use of nonverbal communication	.30 [.19, .40]	.24 [.15, .34]	.08 [.02, .17]	.14 [.02, .30]
Facial expression	.36 [.25, .47]	.49 [.37, .60]	.15 [.06, .26]	.24 [.08, .43]
Nonverbal understanding	.52 [.41, .64]	.42 [.31, .54]	.21 [.10, .34]	.24 [.08, .44]
Teacher report				
Use of nonverbal communication	.47 [.34, .61]	.31 [.18, .44]	.38 [.24, .53]	.15 [.01, .37]
Nonverbal understanding	.72 [.59, .84]	.45 [.31, .60]	.30 [.16, .45]	.28 [.06, .56]
Developing, maintaining, and understanding relationships				
Parent report				
Adjusting behaviour for situation	.29 [.19, .39]	.29 [.20, 40]	.10 [.03, .20]	.09 [.00, .23]
Imaginative play	.40 [.29, .51]	.32 [.22, .43]	.26 [.14, .39]	.14 [.02, .30]
Submissive/ dominating in play	.35 [.25, .46]	.50 [.39, .62]	.28 [.16, .41]	.45 [.24, .66]
Possessive/losing	.58 [.47, .69]	.44 [.33, .55]	.30 [.18, .44]	.35 [.15, .56]
Friendship formation	.72 [.62, .82]	.58 [.47, .69]	.24 [.12, .36]	.40 [.20, .61]
Friendship maintenance	.74 [.63, .83]	.72 [.61, .82]	.35 [.22, .49]	.45 [.24, .66]
Social motivation	.22 [.13, .32]	.19 [.10, .28]	.18 [.08, .30]	.18 [.04, .36]
Consistent companions	.42 [.31, .54]	.38 [.27, .50]	.27 [.15, .41]	.24 [.08, .43]
Diagnostician observations				
Friendship understanding	.47 [.36, .59]	.55 [.44, .67]	.32 [.19, .46]	.24 [.08, .44]
Inclusiveness in play	.15 [.08, .24]	.24 [.13, .35]	.07 [.02, .15]	.02 [.00, .09]
Imaginative/spontaneity in play	.28 [.35, .61]	.26 [.13, .40]	.09 [.01, .22]	.19 [.01, .45]
Teacher report				
Friendship formation	.69 [.56, .81]	.51 [.36, .65]	.40 [.26, .56]	.33 [.09, .59]
Friendship maintenance	.70 [.57, .82]	.58 [.44, .71]	.40 [.25, .56]	.42 [.16, .69]
Stereotypical and repetitive behaviour				
Parent report				
Motor stereotypies	.40 [.29, .52]	.42 [.31, .53]	.26 [.14, .38]	.14 [.02, .30]
Toe walking	.15 [.08, .24]	.19 [.10, .28]	.18 [.08, .30]	.04 [.00, .14]
Flapping	.16 [.08, .24]	.21 [.12, .30]	.10 [.03, .20]	.09 [.00, .23]
Spinning	.06 [.02, .12]	.12 [.05, .20]	.07 [.01, .15]	.03 [.00, .12]

	Group: Proportion of total [$HDI_{80\%}$]			
	ASD		Non-ASD	
Behavioural Category	Male	Female	Male	Female
Gross motor mannerism	.12 [.05, .20]	.17 [.09, .26]	.14 [.05, .25]	.18 [.04, .36]
Rocking/ jumping	.27 [.18, .38]	.20 [.12, .30]	.10 [.03, .20]	.18 [.04, .35]
Rigidity	.11 [.05, .18]	.18 [.10, .28]	.08 [.02, .16]	.09 [.00, .22]
Hand mannerisms	.21 [.12, .31]	.16 [.08, .24]	.17 [.17, .28]	.19 [.05, .37]
Self-injurious	.25 [.15, .35]	.22 [.13, .31]	.13 [.04, .23]	.14 [.02, .30]
Repetitive body use	.05 [.01, .10]	.11 [.04, .18]	.12 [.04, .22]	.18 [.05, .37]
Speech/language	.55 [.44, .66]	.44 [.32, .55]	.21 [.10, .33]	.00 [.00, .06]
Echolalia	.18 [.10, .27]	.11 [.05, .19]	.11 [.04, .21]	.03 [.00, .13]
Third person referencing	.04 [.01, .09]	.03 [.01, .09]	.02 [.00, .06]	.07 [.00, .19]
Neologisms	.21 [.12, .31]	.24 [.15, .34]	.08 [.02, .16]	.29 [.11, .49]
Pronoun reversal	.09 [.04, .16]	.12 [.05, .19]	.14 [.05, .24]	.04 [.00, .14]
Repetitive speech	.53 [.42, .65]	.41 [.30, .52]	.41 [.27, .55]	.14 [.02, .30]
Accents	.12 [.06, .20]	.29 [.19, .40]	.04 [.00, .10]	.14 [.02, .30]
Unusual noises	.47 [.36, .59]	.30 [.20, .41]	.19 [.09, .31]	.19 [.05, .37]
Talking to self	.02 [.00, .05]	.05 [.00, .10]	.02 [.00, .06]	.08 [.00, .21]
Odd prosody	.31 [.21, .42]	.37 [.21, .42]	.34 [.21, .48]	.04 [.00, .15]
Object use	.26 [.17, .37]	.32 [.21, .43]	.11 [.04, .21]	.08 [.00, .21]
Lining up	.44 [.32, .55]	.33 [.22, .44]	.43 [.29, .58]	.35 [.20, .52]
Grouping	.18 [.10, .27]	.25 [.16, .35]	.17 [.07, .28]	.24 [.08, .43]
Spinning/flicking/pushing	.25 [.15, .35]	.16 [.08, .24]	.04 [.00, .10]	.04 [.00, .14]
Repetitive play	.09 [.03, .16]	.13 [.06, .21]	.04 [.01, .11]	.09 [.01, .24]
Deconstruction	.24 [.15, .34]	.08 [.03, .15]	.05 [.00, .03]	.03 [.00, .13]
Diagnostician observations				
Stereotypical movement	.11 [.05, .18]	.09 [.03, .16]	.07 [.02, .16]	.03 [.00, .14]
Stereotypical speech/ language	.19 [.11, .29]	.20 [.12, .30]	.08 [.02, .16]	.08 [.00, .22]
Teacher report				
Stereotypical movement	.27 [.14, .42]	.11 [.03, .21]	.13 [.04, .24]	.12 [.00, .32]
Stereotypical speech/ language	.39 [.25, .53]	.19 [.08, .31]	.21 [.09, .35]	.07 [.00, .24]
Stereotypical object use	.15 [.05, .28]	.00 [.00, .03]	.05 [.00, .14]	.04 [.00, .18]
Insistence on sameness, routines and rituals				
Parent report				
Distress at change	.72 [.61, .81]	.65 [.54, .76]	.26 [.14, .39]	.49 [.28, .71]
Routine adherence	.50 [.39, .61]	.55 [.44, .66]	.19 [.09, .31]	.40 [.19, .61]
Task switching/transitioning	.42 [.31, .53]	.47 [.36, .59]	.15 [.06, .26]	.19 [.05, .37]

	Group: Proportion of total [$HDI_{80\%}$]			
	ASD		Non-ASD	
Behavioural Category	Male	Female	Male	Female
Cognitive rigidity	.72 [.61, .81]	.64 [.53, .75]	.19 [.09, .31]	.34 [.15, .54]
Diagnostician observations				
Routine adherence	.03 [.00, .08]	.01 [.00, .05]	.02 [.00, .06]	.01 [.00, .06]
Cognitive rigidity	.17 [.10, .26]	.26 [.17, .37]	.06 [.01, .13]	.09 [.00, .22]
Teacher report				
Distress at change	.47 [.33, .61]	.34 [.21, .48]	.30 [.16, .44]	.24 [.04, .48]
Restricted interests				
Specific program/character	.27 [.18, .37]	.42 [.31, .53]	.37 [.23, .51]	.24 [.08, .43]
Random objects	.26 [.35, .57]	.40 [.29, .51]	.17 [.07, .28]	.39 [.20, .61]
Vehicles	.35 [.24, .46]	.01 [.00, .04]	.19 [.09, .31]	.13 [.02, .29]
Toys	.36 [.26, .47]	.39 [.28, .50]	.34 [.21, .48]	.19 [.05, .37]
Screens	.52 [.40, .63]	.17 [.09, .26]	.25 [.14, .38]	.03 [.00, .13]
Animals	.22 [.14, .32]	.28 [.18, .39]	.12 [.04, .23]	.04 [.00, .15]
Systems	.13 [.16, .21]	.11 [.05, .19]	.02 [.00, .07]	.03 [.00, .13]
Craft	.11 [.04, .18]	.24 [.15, .34]	.06 [.01, .13]	.13 [.02, .29]
Sport/activity	.13 [.06, .22]	.12 [.06, .20]	.14 [.06, .25]	.23 [.07, .42]
People	.05 [.01, .10]	.12 [.05, .20]	.00 [.00, .02]	.03 [.00, .12]
Diagnostic observations				
Restricted interest	.00 [.00, .05]	.26 [.17, .37]	.12 [04, .22]	.41 [.30, .53]
Teacher report				
Restricted interest	.61 [.47, .75]	.31 [.18, .45]	.28 [.15, .43]	.15 [.01, .36]
Sensory behaviours				
Parent report				
Auditory: seeking	.13 [.06, .21]	.11 [.05, .19]	.08 [.02, .17]	.09 [.00, .22]
Auditory: avoiding	.71 [.60, .80]	.63 [.52, .74]	.28 [.16, .41]	.40 [.20, .61]
Tactile: seeking	.50 [.39, .61]	.47 [.36, .58]	.23 [.12, .36]	.14 [.02, .30]
Tactile: avoiding	.58 [.47, .70]	.69 [.58, .79]	.30 [.18, .44]	.40 [.20, .61]
Olfactory: seeking	.14 [.07, .23]	.15 [.08, .24]	.06 [.01, .13]	.04 [.00, .14]
Olfactory: avoiding	.21 [.13, .31]	.28 [.18, .38]	.04 [.00, .10]	.09 [.00, .22]
Oral: seeking	.40 [.29, .51]	.52 [.41, .64]	.21 [.10, .33]	.24 [.07, .43]
Oral: avoiding	.39 [.28, .50]	.31 [.21, .42]	.10 [.03, .20]	.24 [.08, .43]
Visual: seeking	.25 [.15, .35]	.29 [.19, .39]	.10 [.03, .19]	.23 [.07, .43]
Visual: avoiding	.16 [.09, .25]	.12 [.05, .20]	.02 [.00, .07]	.18 [.04, .36]

	Group: Proportion of total [$HDI_{80\%}$]			
	ASD		Non-ASD	
Behavioural Category	Male	Female	Male	Female
Diagnostic observations				
Sensory behaviours	.30 [.20, .41]	.30 [.20, .40]	.14 [.06, .25]	.19 [.05, .37]
Teacher report				
Sensory behaviours	.33 [.20, .47]	.27 [.15, .40]	.14 [.05, .26]	.15 [.01, .37]

Table G.2

Results of Logistic Regressions for Behaviours Excluded Due to Low Frequencies

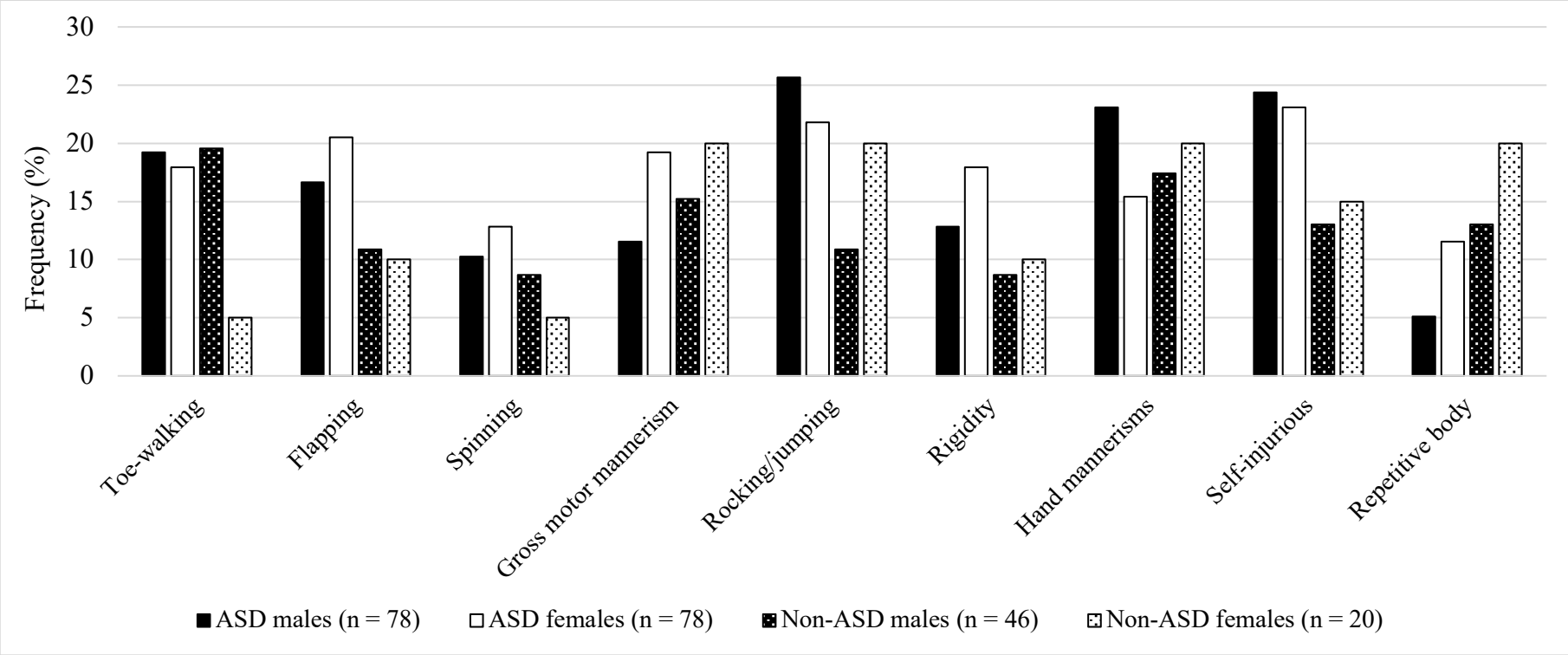
	Effect of Ax. Result		Effect of Sex/Gender		Ax. Result × Sex/Gender Interaction		Prop. Diff. M F (Y-N)
Behavioural Category	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)	
Parent report							
Stereotypical behaviour							
Mouth mannerisms	0.10 [-0.88, 1.05]	.50	0.76 [-0.23, 1.73]	.82	0.40 [-1.44, 2.34]	.58	.00 .01
Restricted interests							
Different places/times	0.09 [-2.36, 2.50]	.50	**5.64 [2.34, 8.77]**	**1.00**	0.89 [-3.85, 5.40]	.59	.01 .00
Self-presentation	**4.78 [1.28, 8.14]**	**.99**	-1.90 [-4.54, 0.59]	-.83	-1.34 [-6.13, 3.31]	-.63	.01 .11
Diagnostic observations							
Stereotypical object use	**2.74 [0.78, 4.69]**	**.98**	**2.32 [0.40, 4.17]**	**.96**	-1.11 [-4.62, 2.54]	-.65	.10 .24
Task switching/ transitioning	**5.59 [2.40, 8.60]**	**1.00**	-0.35 [-2.66, 1.94]	-.56	-0.02 [-4.55, 4.48]	-.49	.05 .08
Teacher report							
Routine adherence	**2.29 [0.49, 4.02]**	**.98**	1.34 [-0.38, 3.07]	.85	-2.73 [-6.06, 0.67]	-.87	.04 .08

Note. Positive LOR (ASD assessment result) = greater probability of being reported if the assessment result was positive for ASD; Positive LOR (sex/gender) = greater probability of being reported for males. Difference in proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). Differences in boldface indicate the $HDI_{80\%}$ lay entirely outside of the ROPE. $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio.

Appendix H: Frequency of Stereotypical Behaviours (Criterion B1) and Restricted Interests (Criterion B3) by Sex/Gender and Assessment Result (Study 2b)[67]

Figure H.1

Frequency of Parent Reported Motor Stereotypies (Criterion B1) by Sex/Gender and Assessment Result



[67] These figures were derived from the raw proportions data rather than the models. Therefore, HDIs $_{(80\%)}$ are not included.

Figure H.2

Frequency of Parent Reported Speech Stereotypies (Criterion B1) by Sex/Gender and Assessment Result

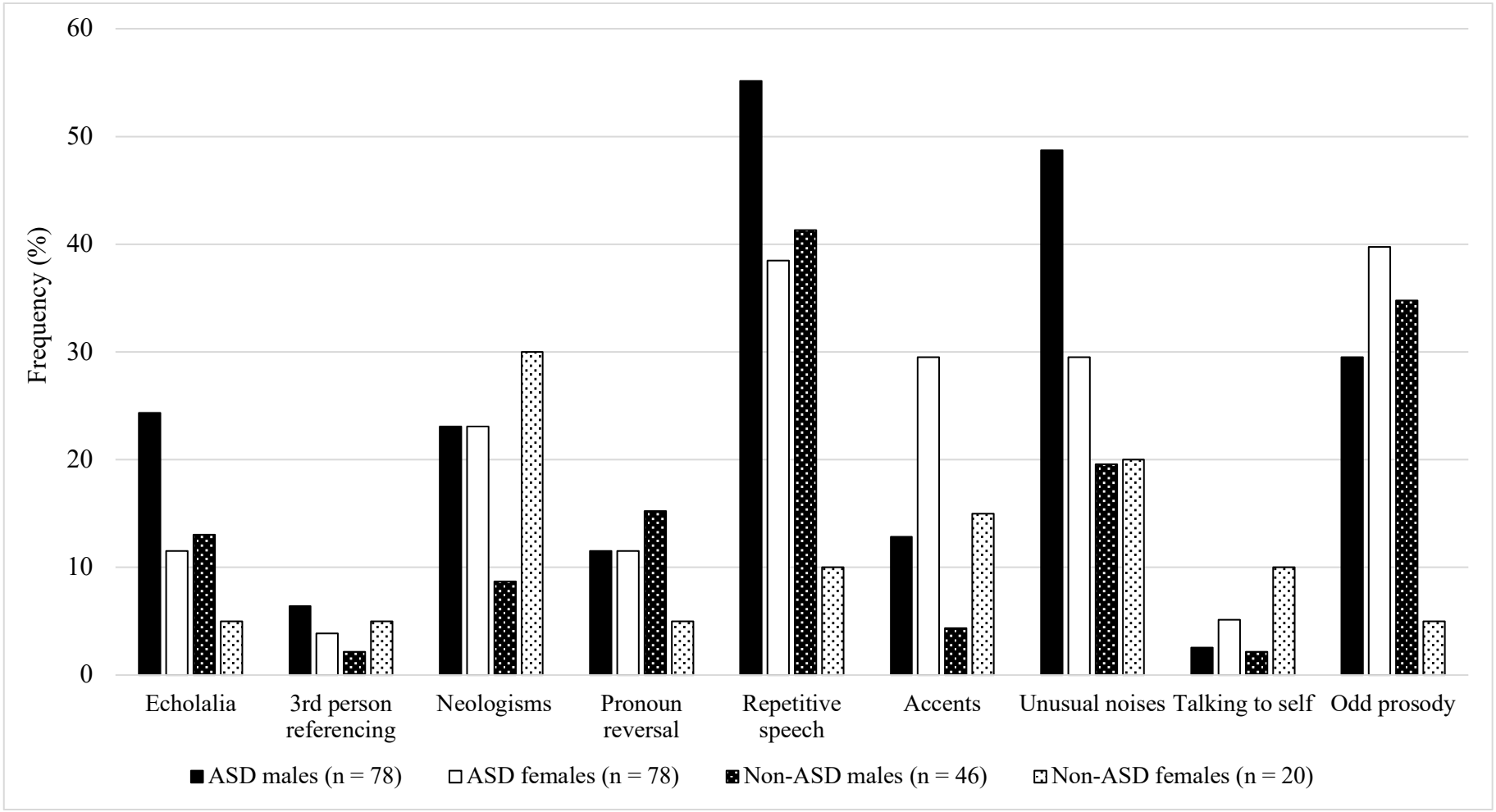


Figure H.3

Frequency of Parent Reported Object Use Stereotypies (Criterion B1) by Sex/Gender and Assessment Result

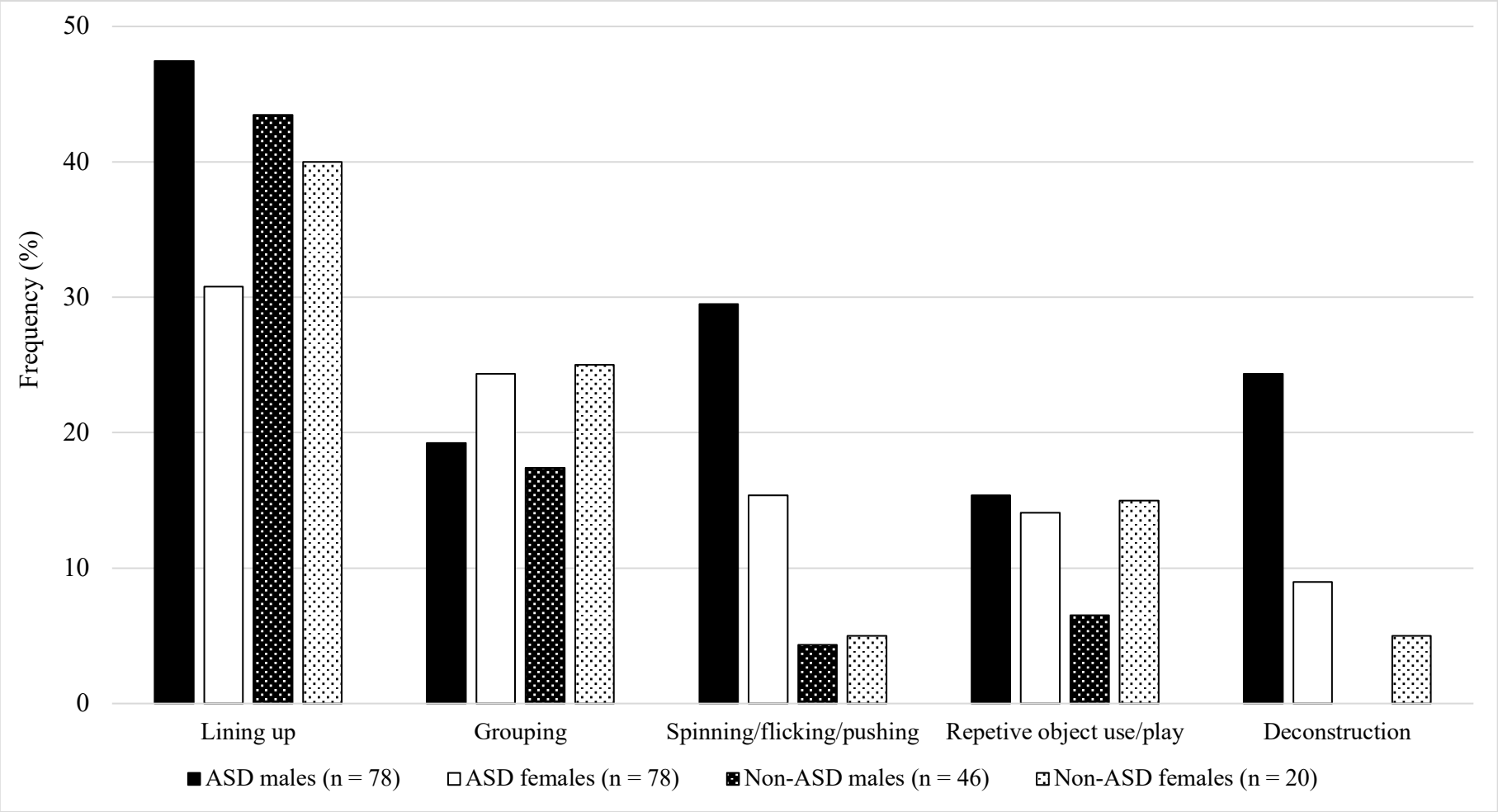
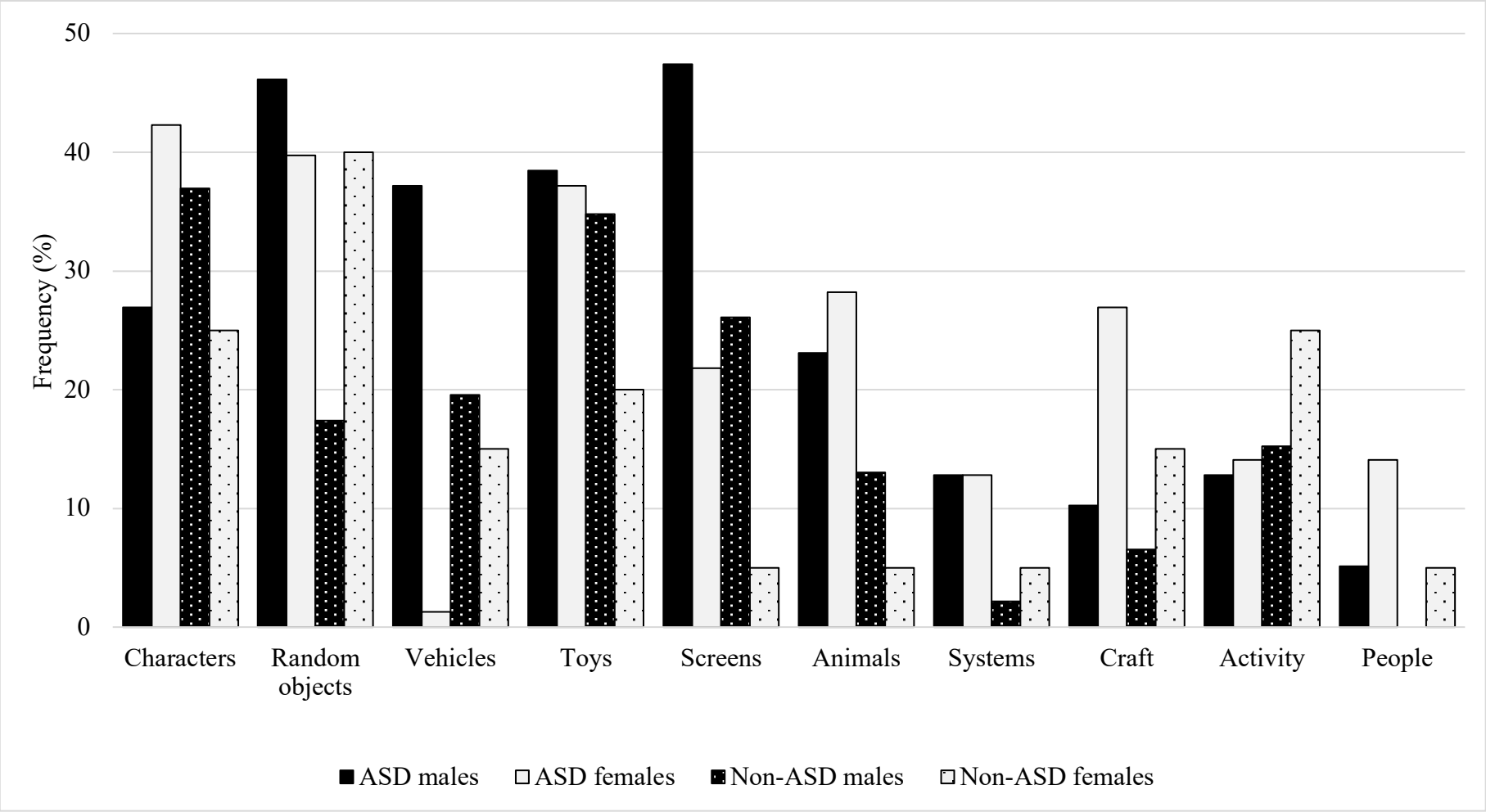


Figure H.4

Frequency of Parent Reported Restricted Interests (Criterion B3) by Sex/Gender and Assessment Result



Appendix I: Meaningful Assessment Result by Sex/Gender Interactions (Study 2b)

Table I

Summary of Behavioural Categories with Meaningful Interactions ($HDI_{80\%}$ Entirely Outside ROPE)

Behavioural category	Source	Ax. Result × Sex/Gender Interaction		Prop. Diff. (Y-N)		ASD: M-F		Non-ASD: M-F	
		LOR [$HDI_{80\%}$]	$P_{(meaning.)}$	M	F	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Social communication									
Content of conversation	Diagnost. obs.	**-2.24 [-3.69, -0.77]**	**-.99**	.28	.53	-0.10 [-0.55, 0.34]	-.51	**2.16 [0.75, 3.54]**	**.99 ***
Friendship formation	Parent report	**1.39 [0.53, 2.27]**	**.97**	.48	.18	**0.62 [0.16, 1.08]**	**.93 ***	-0.76 [-1.49, -0.00]	-.87
Inclusiveness in play	Diagnost. obs.	**-1.85 [-0.18, -3.46]**	**-.93**	.08	.22	-0.55 [-1.11, -0.02]	-.85 *	1.29 [-0.23, 2.84]	.86 *
Imagination/ spont. in play	Diagnost. obs.	**1.88 [3.47, 0.29]**	**.92**	.39	.07	**0.99 [0.39, 1.60]**	**.97 ***	-0.89 [-2.32, 0.63]	-.76
Stereotypical behaviour									
Toe walking	Parent report	**-1.96 [-3.48, -0.38]**	**-.96**	-.03	.15	-0.22 [-0.77, 0.36]	-.61	**1.74 [0.27, 3.15]**	**.96 ***
Speech/ language	Parent report	**-3.63 [-6.29, -0.80]**	**-.99**	.34	.44	0.47 [0.02, 0.91]	.86	**4.05 [1.26, 6.74]**	**1.00**
Third person referencing	Parent report	**1.77 [0.26, 3.67]**	**.87**	.02	-.04	0.19 [-0.86, 1.20]	.54	-1.55 [-3.26, 0.11]	-.88 *
Neologisms	Parent report	**1.37 [0.28, 2.42]**	**.94**	.13	-.05	-0.17 [-0.68, 0.35]	-.57	**-1.54 [-2.46, -0.59]**	**-.98 ***
Odd prosody	Parent report	**-2.75 [-4.15, -1.26]**	**-1.00**	-.03	.33	-0.25 [-0.73, 0.20]	-.66	**2.50 [1.10, 3.87]**	**1.00 ***
Deconstruction	Parent report	**4.11 [0.88, 7.20]**	**.97**	.24	.05	-0.39 [-1.02, 0.23]	-.73	-0.83 [-2.03, 0.36]	-.79
Stereotypical object use	Teacher report	**4.34 [1.09, 7.72]**	**.97**	.10	-.04	**4.54 [1.66, 7.35]**	**1.00 ***	0.29 [-1.62, 2.13]	.55

Behavioural category	Source	Ax. Result × Sex/Gender Interaction		Prop. Diff. (Y-N)		ASD: M-F		Non-ASD: M-F	
		LOR [$HDI_{80\%}$]	$P_{(meaning.)}$	M	F	LOR [$HDI_{80\%}$]	P (meaning.)	LOR [$HDI_{80\%}$]	P (meaning.)
Routines and rituals									
Distress at change	Parent report	**1.34 [0.48, 2.21]**	**.97**	.46	.16	0.30 [-0.17, 0.76]	.70	**-1.04 [-1.78, -0.32]**	**-.95 ***
Cognitive rigidity	Parent report	**1.12 [0.23, 2.04]**	**.93**	.53	.30	0.35 [-0.13, 0.80]	.75	-0.77 [-1.55, 0.00]	-.86 *
Restricted interests									
Specific program/ character	Parent report	**-1.25 [-2.17, -0.36]**	-.95	-.01	.18	**-0.65 [-1.10, -0.18]**	**-.94**	0.59 [-0.20, 1.36]	.80
Random objects	Parent report	**1.43 [0.56, 2.33]**	**.97**	.29	.00	0.27 [-0.18, 0.70]	.69	**-1.16 [-1.92, -0.37]**	**-.96 ***
Vehicles	Parent report	**3.38 [1.73, 5.02]**	**1.00**	.15	-.12	**3.77 [2.41, 5.00]**	**1.00 ***	0.46 [-0.51, 1.46]	.68
Sensory behaviour									
Oral: avoiding	Parent report	**1.35 [0.32, 2.36]**	**.94**	.29	.07	0.34 [-0.10, 0.82]	.75	-0.99 [-1.93, -0.10]	-.89 *
Visual: avoiding	Parent report	**2.88 [1.21, 4.45]**	**.99**	.14	-.06	0.39 [-0.23, 1.01]	.73	**-2.47 [-4.00, -0.99]**	**-.99 ***

Note. Positive interaction LOR = behaviour is more strongly associated with ASD result for males, negative LOR = behaviour is more strongly associated with ASD result for females (underlined). Difference in proportion of children with behaviour reported for males (Yes - No ASD result) and females (Yes - No ASD result). $P_{(meaningful)}$ = probability that the true difference fell outside the ROPE and in the observed direction. LOR = log odds ratio. * Indicates that the sex/gender difference differed substantially between the ASD and non-ASD group

Appendix J: Recruitment Flyer (Study 3)

Image removed due to copyright restrictions.

Diagnostic Assessment of Autism Spectrum Disorder

Do you conduct diagnostic assessments for Autism Spectrum Disorder?

Please consider our invitation to participate in an online questionnaire about challenges associated with these diagnostic assessments and your experiences working with females with ASD.

The questionnaire will take approximately 45 minutes.
Upon completion, you will be asked to provide an email address if you would like to receive a $75 e-gift card for your time.
Participation is entirely voluntary.

To participate, please visit
https://qualtrics.flinders.edu.au/jfe/form/SV_agz9ygveZX8Chud
Password: autism

If you have any questions or feedback about the study, please contact Joanna Tsirgiotis at joanna.tsirgiotis@flinders.edu.au

Flinders UNIVERSITY

Appendix K: Diagnostician Questionnaire (Study 3)

Thank you for considering our invitation to complete this questionnaire.

Q1.1. Are you a clinician working with individuals with autism spectrum disorder (ASD)?

- ☐ Yes
- ☐ No

Q1.2. Are you a delegate of Professor Robyn Young's workshop ***'Autism spectrum disorder: The female presentation and the link with eating disorders'*** to be held at the APS College of Clinical Psychologists conference?

- ☐ Yes
- ☐ No

LETTER OF INTRODUCTION

Dear Sir/Madam,
I hold the position of Professor in the College of Education, Psychology and Social Work at Flinders University. This letter is to introduce Joanna Tsirgiotis who is a PhD (Clinical Psychology) student.

Joanna is undertaking research leading to the production of a thesis and other publications on the subject of diagnostic assessment of autism spectrum disorder. Joanna would like to invite you to participate in this project by completing a questionnaire which covers certain aspects of this topic, including your experiences working with females with ASD and challenges in assessment.

Please be assured that any information provided will be treated in the strictest confidence and none of the participants will be individually identifiable in the resulting thesis, report or other publications. You are, of course, entirely free to discontinue your participation at any time or to decline to answer particular questions. This project has been approved by the Social and Behavioural Research Ethics Committee (SBREC; project number 8302).

We hope that you will accept this invitation to be involved. Should you do so, we ask that you do not divulge details of the content of the questionnaire to others. Any enquiries you may have concerning this project should be directed to me at the address given above or by telephone on (08 8201 5194) or email (robyn.young@flinders.edu.au).

Thank you for your attention and assistance.

Yours sincerely,
Professor Robyn Young
Professor of Psychology
College of Education, Psychology and Social Work, Discipline of Psychology

INFORMATION SHEET

Title: Diagnostic Assessment of Autism Spectrum Disorder

Researcher:
Joanna Tsirgiotis
College of Education, Psychology and Social Work Flinders University
Tel: 8271 2370

Supervisors:
Professor Robyn Young; Associate Professor Nathan Weber
College of Education, Psychology and Social Work
Flinders University
Tel: 8201 5104; 8201 2968

Description and purpose of the study

This project will investigate challenges associated with diagnostic assessment of autism spectrum disorder. This project is supported by Flinders University, College of Education, Psychology and Social Work.

What will I be asked to do?

Participation in this study is voluntary. That is, it is at your discretion as to whether you wish to complete this questionnaire and whether your responses are used in this research.
The questionnaire involves two case studies and follow-up questions. The questionnaire will take approximately 45 minutes to complete. Upon completion of the questionnaire, you will be given a $75 e-gift card for your time and participation (redeemable at a variety of outlets).

What benefit will I gain from being involved in this study?

There are no direct benefits to your participation, however your involvement will help advance our understanding around ASD assessment and diagnosis.

Will I be identifiable by being involved in this study?

We do not need your name and you will be anonymous. Your responses will not be linked directly to you. All information and results obtained in this study will be stored in a secure way, with access restricted to relevant researchers. You may choose to provide your email address in order to receive your e-gift card and/or the results of the study.

The provision of your email address may mean that your identity becomes known to the researchers (i.e., if your email address contains your name). However, no identifying information will be published, and your email addresses will be separated from your questionnaire responses.

Are there any risks or discomforts if I am involved?

The investigators do not anticipate any risks resulting from your involvement in this study. If

you have any concerns regarding anticipated or actual risks or discomforts, please feel free to raise them with the investigators.

How do I agree to participate?
Should you volunteer for your responses to go towards this research, you may refuse to answer any questions and you are free to withdraw at any time without effect or consequences. A consent form accompanies this information sheet.

How will I receive feedback?
Details of the study's purpose and aims will be presented at the conclusion of the questionnaire and discussed at the Australian Psychological Society College of Clinical Psychologists 2019 conference for delegates of Professor Robyn Young's workshop. A form debriefing participants of the aims of the study can be found at the conclusion of the questionnaire. Participants will be given the option of requesting that the results of the study be sent to them via email following data analysis.

After October 31st 2019, a link with the results of the study will be made available for those participants who choose not to leave their email address.

Thank you for taking the time to read this information sheet, and we hope that you will accept our invitation to be involved.

This research project has been approved by the Flinders University Social and Behavioural Research Ethics Committee in South Australia (Project number 8302). For queries regarding the ethics approval of this project, or to discuss any concerns or complaints, please contact the Executive Officer of the committee via telephone on +61 8 8201 3116 or email human.researchethics@flinders.edu.au

Q2.2.
I am over the age of 18 years hereby consent to participate as requested above.

i. I have read and understood the information provided above.
ii. Details of procedures and any risks have been explained to my satisfaction.
iii. I understand that:

- I may not directly benefit from taking part in this research.
- Participation is entirely voluntary and I am free to withdraw from the project at any time; and am free to decline to answer particular questions.
- While the information gained in this study will be published as explained, my participation will be anonymous and my individual information will remain confidential.

iv. I understand that only the researchers on this project will have access to my research data and raw results; unless I explicitly provide consent for it to be shared with other parties.
v. I agree not to divulge or share the content of the questionnaire with any other parties.

- ☐ Yes, I consent.
- ☐ No, I do not consent. I wish to complete the questionnaire for professional development, but I do not wish for my responses to be included in this research. I will therefore be ineligible for reimbursement.
- ☐ No, I do not consent to completing the questionnaire nor participation in research.

Q3.1. What is your profession?

- ☐ Psychology
- ☐ Speech Pathology
- ☐ Occupational Therapy
- ☐ Psychiatry
- ☐ Paediatrics (medical)
- ☐ Other (please specify): _____

Q3.2. In which Australian state or territory do you primarily practice?

- ☐ Australian Capital Territory
- ☐ New South Wales Northern Territory
- ☐ Queensland
- ☐ South Australia
- ☐ Tasmania
- ☐ Victoria
- ☐ Western Australia
- ☐ New Zealand

Q3.3. What is your sex?

- ☐ Female
- ☐ Male
- ☐ Other
- ☐ I would prefer not to disclose

Q3.4. For how many years have you worked in a clinical setting with individuals with autism spectrum disorder? _____

Q3.5. Do you conduct assessments for autism spectrum disorder?

- ☐ Yes
- ☐ No
- ☐ I have in the past but not currently

Q3.6. For how many years have you been conducting assessments for autism spectrum disorder? _____

Q3.7. How many of each of the following groups of people do you assess for autism spectrum disorder in an average 6-month period?

	None	1-3	3-6	7-10	11 or more
Boys in early childhood (0-4 years)	☐	☐	☐	☐	☐
Girls in early childhood (0-4 years)	☐	☐	☐	☐	☐
School age boys (5-11 years)	☐	☐	☐	☐	☐
School age girls (5-11 years)	☐	☐	☐	☐	☐
Adolescent boys (12-17 years)	☐	☐	☐	☐	☐
Adolescent girls (12-17 years)	☐	☐	☐	☐	☐
Adult men (18 years or more)	☐	☐	☐	☐	☐
Adult women (18 years or more)	☐	☐	☐	☐	☐

<<Female case study: Grace/Gordon condition>>

Q4.1. Please read the following excerpts carefully. They have been designed to mirror diagnostic reports but do not represent any real people.

You will not be able to go back and change your answers. It might be useful to keep note of which criteria you deem met in order to answer the follow up questions. The criteria will be presented to you in a random order.

We understand that it is not possible to form a thorough diagnostic opinion without interacting with an individual and having additional background information. However, please do your best to answer the questions given the information provided below.

Background information

Grace was referred for assessment by Dr Grey (paediatrician) in response to parental concern about emotional regulation and friendships. Grace lives at home with her mother (Debbie), father (John) and her sister Bianca (15yo). She attends Parkview Grammar and is in year 5. Grace was born premature at 35 weeks and had a birth weight of 2.7kg. She was described as a quiet and placid baby who fed regularly and slept well. Grace never crawled but bum-shuffled at 1 year of age and then walked at 14 months. Her language use was thought to be highly developed from a young age. There is no known family history of autism spectrum disorder, learning disorders or mental health difficulties.

Grace has not received any previous assessment or intervention. Debbie has had concerns about Grace's socioemotional development for some years, but no concerns had ever been raised by the school.

Assessment procedure

Grace was assessed by two diagnosticians (a clinical psychologist and a speech pathologist). John and Debbie were interviewed. Grace was observed and engaged throughout the assessment. Grace's teacher, who has known her for 6 months, completed a teacher questionnaire. This information was considered against the *Diagnostic and Statistical Manual for Mental Disorders – 5th Edition.*

Q5.1.

Criteria A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays.

Criterion A1. Deficits in social-emotional reciprocity; ranging, for example, from abnormal social approach and failure of normal back and forth conversation through reduced sharing of interests, emotions, or affect to failure to initiate or respond to social interaction.

Parent report

It is rare for Grace to initiate an interaction; even with people she knows well. When she does initiate conversation, it may not always be at the best time, such as when her parents are talking to each other. If someone initiates a conversation with Grace, she will happily talk about things that interest her, particularly drawing. She will often bring up drawing or other topics out of the blue during a conversation. Grace will often ask about people's weekends but does not really extend the conversation. If others offer information about something in which she is not interested, she might say "oh" or "okay", but rarely asks questions in return. She is able to maintain some reciprocal conversation (albeit limited), but this will depend on the topic and the person with whom she is interacting. She may appear lost or confused when listening to someone talk for an extended period. Grace gets frustrated when others interrupt her. She can talk for some time about an area of interest even when the other person does not appear to share that interest. However, her mother does not believe these are monologues.

As a young child, Grace would always cling to her mother when people she didn't know were around. She would cry when her mother or father were not nearby. Now, when people visit her home, Grace is quiet but can engage with them to some extent. She will still often look to Debbie for reassurance and will not initiate conversation or be overly responsive when others talk to her. Grace went through a phase around age 6 where she would lie about lots of things (e.g., if she had homework or what she had eaten for lunch). She would apologise if found out but carried on with this behaviour for about a year.

Teacher report

Grace is generally shy but likes to talk about her drawing and animals. She tends to be quiet

in conversation with others and does not dominate them, preferring to ask questions and listen. However, when she is discussing something about which she is passionate, Grace may talk over others. Grace doesn't usually initiate conversation, but she will usually engage when spoken to. Grace never asks for help at school but will accept it if she is offered help. She never asks questions during class.

Observations
Upon meeting the assessors, Grace appeared nervous and teary but quickly relaxed. She did not initiate conversation but enjoyed sharing photos of her drawings and it was not difficult to engage her in reciprocal conversation about this. Reciprocal conversation on other topics was more challenging for her but she responded appropriately to all questions asked of her. Grace was able to ask one or two social questions but only when led to do so. I told her that I had majored in visual art at University and had an interest in drawing cats. She nodded but did not engage further. Grace did not always use social pleasantries such as saying "thank-you" in response to a compliment. Instead, she would look down at her feet or around the room.

Q5.2. Based on this information, do you believe that Criterion A1 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q5.3. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.4. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q5.5.
Criteria A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays.

Criterion A2. Deficits in nonverbal communicative behaviours used for social interaction; ranging for example from poorly integrated- verbal and nonverbal communication, to abnormalities in eye contact and body- language, or deficits in understanding and use of gestures, to total lack of facial expression and nonverbal communication.

Parent report
Grace's facial expression can often be flat, but it is possible for her parents to read her emotions, which are rarely out of context when Grace is interacting with them. They have observed, however, that her facial expressions can appear somewhat exaggerated when she is with peers. She is able to hold her emotions together at school but when she gets home she will be 'set off' by small things (i.e., she will cry and withdraw to her room where she will stay for hours if someone so much as looks at her the wrong way). She becomes quite hyperactive when she is very upset and finds it helpful to run or jump around. Occasionally, Grace will burst into tears in the car afterschool and be unable to articulate what is wrong.

Grace will sometimes stare or look into the distance but will refocus when someone tries to get her attention. Her eye- contact has improved but was avoidant when she was a toddler. Grace does not use a lot of descriptive gestures and prefers to use words but will wave to greet and farewell people. She is able to read obvious facial expressions but may not be able to differentiate between similar ones, such as frustration and annoyance. Grace sometimes struggles to interpret others' body language and may at times assume others are angry or laughing at her when this isn't necessarily the case. If this happens at school, Grace will withdraw. With immediate family members she may ask if they are angry or sad. She is not always convinced by their response.

Grace will attempt to console others who are upset. She becomes very upset if she sees another person crying and will always tell an adult. She will try to offer practical assistance but may not know how to comfort the person and feels awkward doing so. If Grace has hurt herself, she will accept comfort.

As a toddler, Grace did not gesticulate to gain others' attention, but she did wave and nod. Debbie could not recall whether she could follow a point. She was described as a stoic baby who didn't cry often.

Teacher report
Grace's eye contact is usually appropriate but can be a little intense at times. She demonstrates a variety of facial expressions and uses gestures for greetings or describing something. Her teacher has not seen her respond to the emotions of others but believes she would be able to do this if the emotions were obvious. She shows concern for others but appears to bottle up her own emotions. When someone is hurt or sad, she may become so overwhelmed that the attention is redirected to her rather than the person who is injured or hurt.

Observation
Grace's affect was generally flat, but she would smile when talking about something she liked. Her eye contact could fall into a stare at times. Her use of nonverbal gestures was limited but she demonstrated a wide vocabulary and sometimes used her hands to emphasise her point. Grace responded empathetically when the assessor feigned accidental injury and

asked her several times later during the assessment if she was alright. Grace reported that she is good at recognising the emotions of others but is not always comfortable responding and usually doesn't know what to do.

Q5.6. Based on this information, do you believe that Criterion A2 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q5.7. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.8. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q5.9.
Criteria A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays.

Criterion A3. Deficits in developing, maintaining and understanding relationships, ranging for example from difficulties adjusting behaviour to suit different social contexts to difficulties in sharing imaginative play and in making friends to an apparent absence of interest in peers.

Parent report
Grace is motivated to make friends and be social and talks to Debbie about her peers. She has a friend (Imogen) in the year below her who also attends art classes with Grace. They spend most of their time drawing together. She has been friends with Imogen for about 1-2 months. Grace much prefers to spend one on one time with friends and can get jealous if others join the friendship.

Grace is easily led by others and may not always identify if other children are being mean to her. She may 'latch' on to new friends and be quite intense. Additionally, Grace has difficulty repairing relationships when conflicts arise, and as a result, she will move from being friends with one child to another over a semester. Grace has strong views about who she likes at school and who she does not like. She will not initiate any interaction with the latter group but is 'quite social' with people she has known for a while.

Grace can require guidance in supporting her friends. For instance, when her Imogen's cat passed away, Grace did not know how to comfort her friend but recognised the need to do so and asked Debbie what to do. In a group setting, Grace tends to become very quiet. She seems to enjoy birthday parties but is exhausted when she comes home and may cry or withdraw for a couple of hours. At a party, she may retreat to the bathroom for 10 minutes at a time but generally looks happy to be there. She often forgets other children's names and instead describes them by their appearance.

As a younger child, Grace was extremely imaginative in solo play with figurines and animal toys. She would direct her sister when playing together. She would often replicate scenes from television in which people were talking to each other. Grace's play sometimes had dark themes, such as characters dying or going missing. At preschool, she would usually play alongside other children but was happy for another child to join her. She was upset if children did not play a game her way.

Teacher report
Grace has one consistent friendship at school but finds it difficult to incorporate others and has difficulty interacting in groups. She can also become jealous when others interact with her friend. This results in relationships with others, or at least potential relationships, breaking down. Conflict greatly upsets Grace, and she can 'shut down' and withdraw when this happens. Grace is respectful to teachers and tries her best at school. She comes across as quite shy.

Observation
Grace reported that she usually hangs out with Imogen (a year below her) during break time at school and that they generally walk around together. If Imogen is away, Grace said she would probably read on her own. She said she prefers small groups and that she can find friendships tricky at times but struggled to articulate why. She does not have play dates but may be invited to some birthday parties. She reported that she only sometimes feels lonely.

Q5.10. Based on this information, do you believe Criterion A3 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q5.11. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.12. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q5.13.
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.

Criterion B1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).

Parent report
Grace may click her pen repetitively when nervous but is able to stop when asked. She is fidgety and struggles to sit still for a long time. When excited, Grace will talk rapidly. She was somewhat echolalic as a small child, but this had improved by the time she started school. When Grace is very upset, she will bang her head with her hand. It is quite difficult for Debbie to get her to stop this and she will require some time to deescalate from this state. She will sometimes do this when in a public place without someone that she knows.

Grace has never grouped or stacked objects and no motor mannerisms were reported. She has never spoken in an accent nor engaged in any unusual vocalisations. She may engage in repetitive questioning, but this mostly relates to changes in routine. No categorising or lining up was reported. Indeed, her bedroom is very messy, and she can never remember where she has put things.

As a toddler, Grace would twinkle her fingers when excited and shake objects that made sounds. She occasionally held objects close to her eyes, but this all stopped by the time she started preschool.

Teacher report
Grace repetitively clicks her pen at school. If excited, she can wave her arms about. If the routine is altered, Grace will ask questions repetitively.

Observations
No unusual behaviours in this domain were observed. Grace sometimes twirled her hair, but this was deemed more a sensory seeking behaviour.

Q5.14. Based on this information, do you believe that Criterion B1 is met?
- ☐ Yes
- ☐ No
- ☐ Partially

Q5.15. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.16. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q5.17.
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.

Criterion B2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behaviour (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).

Parent report
Grace is perfectionistic and highly self-critical. She likes to follow rules but can cope if they are broken by others. She can be very stubborn if she does not want to do something.
Grace does not easily adapt to changes in her daily routine (e.g., going to Grandma's house on the wrong day) and this has been the case for some time. If a change occurs, she will ask lots of questions. She has a bedtime routine and will have a lot of difficulty sleeping if this is disrupted. However, new environments are not problematic. She attends school camps and enjoys holidays. Debbie speculated that Grace manages these situations because familiar people are with her. She is generally anxious around unfamiliar people but is not bothered when a family friend from overseas stays at the house. Grace does not like relief teachers but is able to cope, although she may not be able to concentrate as well.

Grace can struggle with transitions if it is an activity she likes. Debbie may have to ask Grace several times to stop drawing and come for dinner. She is always anxious when starting new school years. When she was younger (2-4 years old), Grace would frequently flit from one activity to another.

Grace is fussy about food presentation (e.g., foods must be deconstructed and elements must be grouped in a certain way). She will refuse to eat food that has not been prepared the way that she wants it. She will insist that half her plate remains empty when being served, and never completely finishes a meal. When younger, Grace had to have a certain set of crockery but no longer insists on this.

Grace must have her hair tied back to a particular point and must say goodbye to the dog before leaving the house. Neither John nor Debbie could think of any other rituals that Grace has.

Teacher report
Grace seems to have no difficulty with change in the classroom. She does prefer to sit in the same seat but will move when asked. She is able to cope with relief teachers but has commented that this makes the classroom different and that she doesn't like it. If changes occur, she may ask several questions about the change and why it was occurring.

Observation
No routines or rituals were noted during the assessment. However, Grace came across as somewhat pedantic with dates and times and would correct her mother if she considered any dates were inaccurate. She had set ideas about certain topics that were difficult to shift.

Q5.18. Based on this information, do you believe that Criterion B2 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q5.19. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.20. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q5.21.
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.

Criterion B3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).

Parent report
Grace's favourite hobby is drawing and she can often be found drawing various animals or people. She will often doodle at school and she says that it helps her concentrate, especially when the teacher is talking. A good amount of Grace's conversation will be centred around

drawing and art and she will proudly share her pictures with others. Debbie was unsure whether she would describe this as an ‘obsession’ or ‘preoccupation’ because she is able to transition away from it (after a few attempts) and has other interests too. For instance, she also loves animals and is hyper-sensitive to anyone being ‘mean’ towards her pets. She has a large collection of shells and rocks that she keeps under her bed and may look at from time to time. If these were disturbed, Grace would be very unhappy.

Imogen, Grace’s friend, shares these interests. Grace can latch onto people like Imogen and likes to tell Debbie everything about these children.

When Grace was younger, she was ‘obsessed’ with the Wiggles. She would collect Wiggles figurines, arrange them in order of favourite and be extremely upset if they were moved. This activity consumed most of her free time and she would become distressed if she was asked to transition to another activity. She does not seem to be very interested in anything that is taught at school and avoids homework. Grace struggles to attend to anything for more than five minutes, with the exception of her drawing.

Teacher report
Grace is dedicated to her drawing, but this is not obsession-like.

Observations
Grace was most animated when discussing her art and pets. She often brought drawing into the conversation when talking about other things (e.g., to show a picture that she had drawn of the conversation subject matter). When discussing her behaviour with her mother, Grace continued to draw, largely indifferent to the conversation going on around her.

Q5.22. Based on this information, do you believe that Criterion B3 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q5.23. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.24. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q5.25
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.
Criterion B4. Hyper- or hypo reactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures or excessive smelling or touching of objects).

Parent report
Grace prefers quiet but is not overly bothered by loud noises, although she might comment and avoid some. She is fine with incidental loud noises. Grace may be under-responsive to sounds at times and may not respond to her name, but only when she is very interested in something and focused. This was more of a problem in the past.

Grace will occasionally look at things out of her peripheral vision and notices small details which she will want to investigate. She is bothered by sunlight and needs to wear sunglasses outdoors. Grace is quite sensitive to smells and may comment about unpleasant odours. Grace has been known to chew pens and put cords in her mouth. This stopped when she was about 8 years old. She may now bite her nails. Grace often twirls her hair, and this has persisted even though Debbie has attempted to extinguish this behaviour.

Grace is not bothered by the feeling of clothing or by brushing her teeth or hair but complains that her hair hurts if it is not tied back in a certain position.

Grace's diet is generally restricted and she does not enjoy trying new foods. She cannot stand the taste of spicy foods or the texture of anything slimy.

Grace was thought to have a normal pain threshold. She is sensitive to temperature but may need to be reminded to put on a jumper.

Teacher report
Grace often twirls her hair in class. Although she stops this when asked, she will soon resume. She is not bothered by loud noises.

Observations
Grace twirled her hair often and reported that she prefers quiet environments.

Q5.26. Based on this information, do you believe that Criterion B4 is met?
- ☐ Yes
- ☐ No
- ☐ Partially

Q5.27. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.28. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q5.29.

Additional information:

- Grace can have difficulty falling asleep but has no problem staying asleep.
- She is terrified of insects and will scream inconsolably if a bug lands on her. She will need to leave the area immediately.
- Grace's teacher reported that she is considerably behind her peers across all subjects.

Criterion C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies later in life).

Based on all of the above information, do you believe Criterion C is met?

- Yes
- No
- Partially

Q5.30. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.31. Criterion D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of functioning.

Based on all of the above information, do you believe Criterion D is met?

- Yes
- No
- Partially

Q5.32. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.33. Criterion E. These disturbances are not better explained by intellectual disability or global developmental delay.

Based on all of the above information, do you believe that Criterion E is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q5.34. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.35. Based on your impressions from this case study, would you be inclined to say that this child has ASD?

- ☐ Yes
- ☐ No

Q5.36. Please rate your confidence that this child has ASD.

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q5.37. Please indicate the level support required in each of the following domains.

	NA (not enough criteria are met	Level 1 ("Requiring support")	Level 2 ("Requiring substantial support")	Level 3 ("Requiring very substantial support")
Social communication (criteria A1-A3)	□	□	□	□
Restricted and repetitive patterns of behaviour (criteria B1-B4)	□	□	□	□

Q5.38. Given the information presented, which of the following differential/additional diagnoses would you consider exploring further?

- □ Attention-deficit/hyperactivity disorder
- □ Conduct disorder
- □ Intellectual disability
- □ Generalised anxiety disorder
- □ Language disorder
- □ Obsessive compulsive disorder
- □ Oppositional defiance disorder
- □ Post-traumatic stress disorder
- □ Reactive attachment disorder
- □ Separation anxiety disorder
- □ Social anxiety disorder
- □ Social (pragmatic) communication disorder
- □ None of the above
- □ Other (please specify): _____

Q5.39. Given the information available to you, which of the following diagnoses do you think is most appropriate for this child?

- □ ASD
- □ ASD and other disorder (please specify): _____
- □ Other disorder only (please specify): _____
- □ No diagnosis

Q5.40. How difficult did you find it to arrive at a diagnostic conclusion for this case?

Extremely easy				Neither easy nor difficult						Extremely difficult
0	10	20	30	40	50	60	70	80	90	100

<<Male case study: Bradley/Bridget condition>>

Q6.1. Please read the following excerpts carefully. They have been designed to mirror diagnostic reports but do not represent any real people.

You will not be able to go back and change your answers. It might be useful to keep note of which criteria you deem met in order to answer the follow up questions. The criteria will be presented to you in a random order.

We understand that it is not possible to form a thorough diagnostic opinion without interacting with an individual and having additional background information.

However, please do your best to answer the questions given the information provided below.

Background information
Bradley was referred for assessment by his General Practitioner due to parental concerns about social relationships. Bradley lives at home with his mother (Naomi) and father (Gary) and his brother Michael (14yo). He attends Seaview Grammar and is in year 4.

Bradley was born at term via Caesarean section. He was described as a sweet baby who fed well but had difficulty falling asleep and would sleep only for short bursts. Bradley's motor milestones were mostly met when expected, although his speech was delayed slightly (first words). There is no known family history of autism spectrum disorder, learning disorders or mental health difficulties.

Bradley has not received any previous assessment or intervention. Naomi has had concerns about Bradley's social development for some years.

Assessment procedure
Bradley was assessed by two diagnosticians (a clinical psychologist and a speech pathologist). Gary and Naomi were interviewed, and Bradley was observed and engaged throughout the assessment. Bradley's teacher, who has known him for 6 months, completed a teacher questionnaire. This information was considered against the *Diagnostic and Statistical Manual for Mental Disorders – 5th Edition.*

Q6.2.
Criteria A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays.

Criterion A1. Deficits in social-emotional reciprocity; ranging, for example, from abnormal social approach and failure of normal back and forth conversation through reduced sharing of interests, emotions, or affect to failure to initiate or respond to social interaction.

Parent report
Bradley will occasionally greet other children and will initiate conversation if they are talking about something in which he is interested. From early childhood, he has readily approached strangers and still seems to have no 'stranger danger'. He has approached adults at the park and asked them questions. Naomi has worked to teach him about the risks of doing this. Bradley will answer social questions asked by others, but it is rare for him to ask them.

Bradley can talk for extended periods about his interests and this conversation often seems somewhat one-sided. However, he will stop if asked to do so. He will ask questions of others if he is interested in the conversation but if not, he will not really engage. Bradley will often change the topic of conversation to something that interests him. It is not unusual for Bradley to interrupt others or talk over his parents with a sense of urgency, especially if they are discussing something in which he is interested.

Bradley will share in others' excitement but may become jealous if they have achieved something that he has not. He has a good sense of humour and likes sharing silly jokes but does not understand jokes with double meanings. Bradley often swears at his parents and teachers and can make inappropriate gestures at times. He can swear in public and embarrass Naomi.

As a toddler, Bradley would show, indicate and share objects of interest. He would usually reciprocate a smile, unless occupied by something else. Now, his reciprocation of a smile depends on his mood.

Teacher report
Bradley can sometimes dominate conversations and he will only start a conversation about something he likes. He can sometimes talk over others and interrupt them, but this was more pronounced in the past. Bradley is easily distracted in the classroom and needs regular movement breaks, which help him get back on task. Bradley never asks for help at school but will accept it if offered. He never asks questions during class. He will often call out answers to questions without putting his hand up and this can be disruptive.

Observations
Bradley seemed to struggle with reciprocal conversation. When led to ask a question (e.g., "that's not my favourite toy...") he would say "okay" but then carry on talking about his own preferences. However, he initiated a conversation about an iPad game and asked some questions of the assessor (e.g., "do you have any pets?"). He asked some follow up questions about the assessor's dog but seemed to lose interest after a few exchanges. Bradley did not

talk at length about any particular topic and seemed somewhat disinterested in the assessment process. On several occasions, he asked Naomi when they were going home.

Q6.3. Based on this information, do you believe that Criterion A1 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q6.4. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.5. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q6.6.
Criteria A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays.

Criterion A2. Deficits in nonverbal communicative behaviours used for social interaction; ranging for example from poorly integrated- verbal and nonverbal communication, to abnormalities in eye contact and body- language, or deficits in understanding and use of gestures, to total lack of facial expression and nonverbal communication.

Parent report
Bradley has never had difficulties with eye contact. He is not one to use many gestures and but will make fists or cross his arms when he is cross. When he wishes to engage others, he makes faces to make them laugh, and these faces can be inappropriate at times. It can sometimes be difficult to tell exactly how he is feeling, but it is clear when he is very happy or very angry and this is rarely out of context. Having said this, it is not always immediately clear what has caused his emotion. Bradley can struggle to calm down when his emotions are heightened. When he is angry, Bradley's body stiffens, and he forms fists.

Naomi was unsure as to whether Bradley can read others' body language and emotions. She thought that he could recognise most emotions but may not respond appropriately. For example, if Naomi was sad, he would immediately recognise this, but might make a silly face to cheer her up instead of offering comfort. Naomi felt that this reflected a lack of maturity. He can struggle to read emotions such as subtle frustration or puzzlement. As a younger

child, Bradley did not always identify when his brother was not enjoying a game or persisted despite noticing his brother's lack of enjoyment. Bradley usually stands at an appropriate distance from other people but may occasionally stand too close. This was more pronounced when he was younger.

When Bradley is upset, he will rarely approach Naomi for comfort. When he is *very* upset, it is difficult for anyone to console him and he may throw objects around. He is generally able to provide practical support to others when they are hurt but rarely empathy. Recently, he stopped and help a child who had fallen over by offering him a band-aid.

When Bradley was a toddler, he clapped, nodded and pointed as expected. However, he cried often and was difficult to soothe.

Teacher report
Bradley conveys a range of facial expressions and there are no problems with his eye contact. However, Bradley can be a bit emotionally labile and aggressive when he is very upset. Although this does not happen often, it is disruptive and so he is sent to the office. Bradley can become upset when he does not get his way or is asked to do classwork that doesn't interest him. The teacher was unsure as to whether Bradley could read others' emotions or nonverbal behaviour but commented that Bradley does not usually change his own behaviour in response.

Observation
Bradley's affect was somewhat flat and he did not immediately recognise when the assessor feigned boredom and continued talking. However, at one stage he asked if the assessor was listening. He immediately recognised anger in his mother's voice and asked why she was angry with him, although the anger was directed elsewhere. Bradley appeared able to read gestures and occasionally used some himself to indicate direction or size.

Q6.7. Based on this information, do you believe that Criterion A2 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q6.8. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.9. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q6.10.
Criteria A. Persistent deficits in social communication and social interaction across contexts, not accounted for by general developmental delays.

Criterion A3. Deficits in developing, maintaining and understanding relationships, ranging for example from difficulties adjusting behaviour to suit different social contexts to difficulties in sharing imaginative play and in making friends to an apparent absence of interest in peers.

Parent report
Bradley sometimes seems to lack the degree of genuine interest in his peers that most children would have and so far, has not maintained any friendships for longer than six months. While he believes he has friends, Naomi thinks these are children he just hangs out with at school and plays video games with. At lunchtime, Bradley may sit on his own or play games with Kyle, another boy in his year. Naomi described this child as very sweet-natured. Bradley and Kyle have had play dates and will play video games together but not talk a lot. At school, they play computer games in the library or read. He is consistently invited to a couple of birthday parties each year.

When interacting with other children he can take on a 'silly' role to make others laugh. According to Naomi, other children tend to find him a bit immature, and either intense or disinterested. He can be directive in his play and possessive of his belongings. He is not interested in joining games with other children, unless it is something he likes, and then he can take over.

Whilst at preschool, Bradley enjoyed killing insects and would look for them and bash them with his shoes. He was also entertained by being mean to his brother by stealing his toys or eating his food.

Bradley has difficulty participating in group activities because he can be quite loud and domineering. He is rule oriented during play and but would break them if he was losing a game. As a younger child, Bradley loved playing with Lego and building enormous structures. He also played with Lego figurines. Naomi did not feel that his play was overly imaginative but it varied somewhat from day to day. Bradley was happy for another child to join him in his Lego play, but would not allow them to touch his constructions. Nowadays, much of Bradley's play revolves around video games or Lego.

Teacher report
Bradley struggles socially but has a couple of consistent friends. He does not tend to join in big groups and may choose to wander the yard if he does not feel like playing games with

Kyle. If he is involved in conflict, he can become aggressive and may hit or throw things. His teacher was unsure if he would be able to repair a friendship.

Observation
Bradley reported that he has friends at school and doesn't feel lonely. He said that he gets frustrated when his friends don't play games properly. When asked what makes a good friend, he said that they are, "someone you get on with who likes the same things as you." Bradley allowed the assessor to join him in his Lego construction and was okay with her directing him to build a structure the way she wanted it.

Q6.11. Based on this information, do you believe that Criterion A3 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q6.12. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.13. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q6.14.
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.

Criterion B1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).

Parent report
As a pre-schooler, Bradley used to mix up pronouns. Now, he will say "mum" over and over and asks questions repetitively. He will sometimes hum to himself. This annoys his brother, so Bradley will stop when asked. Bradley has difficulty sitting still and will engage in gross motor movements including rocking, bouncing, and pacing especially when he is on the phone or asking a question.

Bradley likes to organise his figurines by size and lines them up on his window ledge. He engaged in hand flapping as a toddler but this stopped by the time he started school. Naomi

was worried about his echolalia as a younger child but noticed that it stopped when he started school. Bradley loves deconstructing objects but does not put them back together again. He will not organise his Lego by colour (instead leaving it spread all over the floor), but he will categorise his animal figurines.

Teacher report
Bradley has trouble sitting still and will sway in his chair. He also leans back in his chair so that the front legs leave the floor. He is frequently reminded to keep all the chair legs on the floor.

Observation
Bradley wiggled his legs in the chair and at times, tapped the table repetitively. He stopped this when asked to. He did not demonstrate any other unusual movements or use of objects during the course of the assessment.

Q6.15. Based on this information, do you believe that Criterion B1 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q6.16. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.17. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q6.18.
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.

Criterion B2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behaviour (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).

Parent report
Bradley will become very upset if plans change when they involve something that he was looking forward to. In response, he will throw things and become angry. Bradley will notice

and comment if they take a different route in the car or if Naomi cooks with a different ingredient or brand of food. He likes structure and routine and can be a bit uncertain when routines change. He generally copes with this but may be more likely to be 'set off' by other things that go wrong. Bradley has a bedtime routine where he will drink his milk, kiss his mother and father and then read a short story in bed. Although he prefers to this order, he will not be upset by changes in the order. He will be 'out of sorts' if he is prevented from completing this routine (e.g., if sleeping at his grandmother's house).

Bradley likes rules but will break them in order to win a game. However, he is intolerant of other people breaking rules. He likes patterns and will point out patterns that his parents don't notice. When he is asked to, Bradley transitions away from activities well, but he may become distracted and do something else instead. He tends to have difficulty focusing on anything for any more than 10 minutes with the exception of activities he particularly likes.

Teacher report
Bradley is fine with relief teachers but likes to know what he will be doing each day. Class planners are used at school and he is rigid about deviations from these plans. No unusual routines or rituals were noted.

Observation
Bradley developed a routine where he would say "hmm" before answering a question in the assessment. No other ritualistic behaviour was observed.

Q6.19. Based on this information, do you believe that Criterion B2 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q6.20. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.21. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q6.22.
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.

Criterion B3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).

Parent report
Bradley loves Lego and has done since he was a toddler. He has built elaborate castles and other structures with his Lego. He went through a phase where he would watch YouTube videos of children making Lego structures. Naomi commented that Lego time is a good positive reinforcement for Bradley to do his chores. Bradley would scream if his Lego were deconstructed without his permission.

Bradley also loves video games and would play them for hours if allowed. He also likes reading *Horrible Histories* and has a routine of reading one before sleeping each night. He will talk about what he has read, but Naomi did not feel this was excessive.

Bradley has a small collection of rocks which are stored in a special box. Naomi does not think she could throw them away without him becoming very upset.

Teacher report
Bradley loves Lego and will often have some on his desk. He likes to play computer games in the library at lunch time, if it is allowed.

Observation
Upon arrival, Bradley immediately noticed the Lego and asked to play with it. He discussed his interest in history and videogames with the assessor and allowed her to join in the Lego construction.

Q6.23. Based on this information, do you believe that Criterion B3 is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q6.24. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.25. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q6.26.
Criteria B. Restricted, repetitive patterns of behaviour, interests, or activities.

Criterion B4. Hyper- or hypo reactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures or excessive smelling or touching of objects).

Parent report
Bradley used to be scared of fireworks and vacuum cleaners but is no longer bothered by them. He tends to make a lot of noise and likes loud music. However, if there are multiple sounds from different sources occurring at once, Bradley will leave. Bradley will occasionally look at things out of his peripheral vision and notices small details which he will want to investigate. He is bothered by sunlight and needs to wear sunglasses outdoors.

Bradley does not seem to feel the cold and will wear shorts and a T-shirt all year round. He has no issues with the feeling of clothes. Bradley has no sensory difficulties with brushing his hair or teeth but often forgets to do so. He loves to fiddle and will sometimes stroke his mother's hair when he is sitting next to her. Bradley eats a variety of foods of different textures and tastes. He has no abnormal sensitivity to smells but may inappropriately comment about someone's body odour.

Bradley has been known to put Lego and the cord of his hat in his mouth. He used to suck on his sleeves, but this has not occurred for years. Naomi said that Bradley would put anything in his mouth from when he was a toddler to when he started school.

Bradley's pain threshold was thought to be typical.

Teacher report
Bradley is distracted when the classroom noise level is too high. He may chew the ends of pens.

Observations
Bradley was observed to fiddle with Lego when asked questions about his friends.

Q6.27. Based on this information, do you believe that Criterion B4 is met?
- ☐ Yes
- ☐ No
- ☐ Partially

Q6.28. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.29. How severe is the child's impairment in this domain?

Very mild										Very severe
0	10	20	30	40	50	60	70	80	90	100

Q6.30. **Additional information:**

- Bradley has difficulties with both sleep onset and maintenance.
- Bradley's fine and gross motor skills appear slightly delayed.
- Bradley's teacher reported that he is slightly behind his peers in some subjects.

Criterion C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies later in life).

Based on all of the above information, do you believe Criterion C is met?

- Yes
- No
- Partially

Q6.31. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.32. Criterion D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of functioning.

Based on all of the above information, do you believe Criterion D is met?

- Yes
- No
- Partially

Q6.33. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.34. Criterion E. These disturbances are not better explained by intellectual disability or global developmental delay.

Based on all of the above information, do you believe that Criterion E is met?

- ☐ Yes
- ☐ No
- ☐ Partially

Q6.35. How confident are you that this criterion is **met**?

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.34. Based on your impressions from this case study, would you be inclined to say that this child has ASD?

- ☐ Yes
- ☐ No

Q6.35. Please rate your confidence that this child has ASD.

Not at all confident										Extremely confident
0	10	20	30	40	50	60	70	80	90	100

Q6.36. Please indicate the level support required in each of the following domains.

	NA (not enough criteria are met	Level 1 ("Requiring support")	Level 2 ("Requiring substantial support")	Level 3 ("Requiring very substantial support")
Social communication (criteria A1-A3)	□	□	□	□
Restricted and repetitive patterns of behaviour (criteria B1-B4)	□	□	□	□

Q3.37. Given the information presented, which of the following differential/additional diagnoses would you consider exploring further?

- □ Attention-deficit/hyperactivity disorder
- □ Conduct disorder
- □ Intellectual disability
- □ Generalised anxiety disorder
- □ Language disorder
- □ Obsessive compulsive disorder
- □ Oppositional defiance disorder
- □ Post-traumatic stress disorder
- □ Reactive attachment disorder
- □ Separation anxiety disorder
- □ Social anxiety disorder
- □ Social (pragmatic) communication disorder
- □ None of the above
- □ Other (please specify): _____

Q6.38. Given the information available to you, which of the following diagnoses do you think is most appropriate for this child?

- □ ASD
- □ ASD and other disorder (please specify): _____
- □ Other disorder only (please specify): _____
- □ No diagnosis

Q6.39. How difficult did you find it to arrive at a diagnostic conclusion for this case?

Extremely easy				Neither easy nor difficult						Extremely difficult
0	10	20	30	40	50	60	70	80	90	100

Q7.1.
In your clinical experience, are there differences in how males and females with autism spectrum disorder typically present?

- Yes, marked differences
- Yes, moderate differences
- Yes, subtle differences
- No significant differences
- Not sure

Q7.2. How familiar are you with differences between males and females in how autism spectrum disorder may present?

Not at all familiar				Moderately familiar						Very familiar
0	10	20	30	40	50	60	70	80	90	100

Q7.3.
Compared to males, how challenging do you generally find it to form a diagnostic opinion for a female presenting for an autism spectrum disorder assessment?

- Assessing a female is much easier
- Assessing a female is moderately easier
- Assessing a female is slightly easier
- Assessment is equally challenging regardless of sex
- Assessing a male is slightly easier
- Assessing a male is moderately easier
- Assessing a male is much easier

Q7.4. Please rate your level of confidence in accurately diagnosing each of the following groups of people presenting query ASD.

	Not at all confident									Extremely confident
Men	0	10	20	30	40	50	60	70	80	90	100
Women	0	10	20	30	40	50	60	70	80	90	100
Boys	0	10	20	30	40	50	60	70	80	90	100
Girls	0	10	20	30	40	50	60	70	80	90	100

Q7.5. In your opinion, what are three reasons that females with ASD may be under-diagnosed?

1. _____
2. _____
3. _____

Q7.6. In your opinion, what are three features of ASD that may present differently in females?

1. _____
2. _____
3. _____

Q7.7. On average, how much do you rely on each of the following sources of information to form your clinical impression for boys and girls presenting query ASD? (Totals must each equal 100%)

	Parent report	Teacher report	Your own observations	Scores on standardised tools (e.g., SRS)	Total
Boys	□	□	□	□	□
Girls	□	□	□	□	□

Q7.8. In your opinion, what are the most challenging aspects of assessing a female for ASD?

Q7.9. Have you changed anything about your assessment procedures to circumvent such challenges?

- □ Yes
- □ No
- □ Not applicable

Q7.10. What have you changed about your assessment procedures to circumvent such challenges?

Q7.11. In your experience, compared to boys with ASD, girls with ASD are:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree
Better able to camouflage difficulties	☐	☐	☐	☐	☐	☐
More highly motivated to form and maintain relationships	☐	☐	☐	☐	☐	☐
More likely to be disruptive at school	☐	☐	☐	☐	☐	☐
More likely to have social difficulties that only emerge later in life	☐	☐	☐	☐	☐	☐
More likely to present with internalising difficulties	☐	☐	☐	☐	☐	☐
More likely to present with externalising difficulties	☐	☐	☐	☐	☐	☐
More likely to present differently at home compared to school	☐	☐	☐	☐	☐	☐
More likely to be described as ‘shy’ or ‘quirky’ instead of having ASD	☐	☐	☐	☐	☐	☐
More likely to have difficulties with anxiety	☐	☐	☐	☐	☐	☐
More likely to have difficulties with hyperactivity	☐	☐	☐	☐	☐	☐

<<Redirection to separate questionnaire>>

Thank you for completing this questionnaire.

To claim your reimbursement, please enter your email address below. Your email address will be separated from your responses.

Please ensure it is entered **correctly** *and that you check your spam folder periodically.*

Please allow a few weeks for your gift card to be sent.
If you do not wish to leave your email address, you can view the results of the study via the following link from October 31st, 2019.

https://osf.io/2ywck/

Please check the following box if you would like the results of the study to be sent to you via the above email address following data analysis.

- ☐ I would like to receive the results

DEBRIEF OF STUDY AIMS

Researcher
Joanna Tsirgiotis

Supervisors
Professor Robyn Young and Associate Professor Nathan Weber

Diagnostic Assessment of Autism Spectrum Disorder

Part of the project
Sex differences in the Presentation of Autism Spectrum Disorder:
Diagnostic and Clinical Implications.

Thank you for your participation in this project. The purpose of this study was to examine potential challenges in ASD diagnosis associated with gender expectations and experiences of the 'female phenotype' of ASD.

During the questionnaire, you were given two case studies: one, a presentation reflective of that of many boys, and the other, a presentation reflective of that of many girls. Within each of these, the child was randomly allocated a boy's name or girl's name but was otherwise identical across conditions. We were interested to see whether the child's sex influenced whether each ASD criterion was deemed met, your confidence in these decisions and how these related to your experience in assessing girls for ASD. We were also interested in your experiences of assessing females for ASD. The table below reflects the study design and its conditions.

	Male sex condition	**Female sex condition**
Case study 1 'Male' presentation	'Bradley' Condition 1A	'Bridget' Condition 1B
Case study 2 'Female' presentation	'Gordon' Condition 2A	'Grace' Condition 2B

The specific purpose of the study was concealed at the outset as we suspected this knowledge may have influenced results. **We would appreciate it if you would refrain from discussing the specific aims and design of this study with possible future participants.**

If you have any further questions or queries regarding this research, please contact Joanna Tsirgiotis (joanna.tsirgiotis@flinders.edu.au), Flinders University, College of Education, Psychology & Social Work.

Once again, thank you for your participation in our research.

Appendix L: Supplementary Tables and Figures for Case Study Experiment (Study 3)

Table L.1

Diagnostician Endorsement of ASD Criteria, Confidence and Severity Ratings

	Female presentation case study		Male presentation case study	
Criterion	Female condition: *Grace*	Male condition: *Gordon*	Male condition: *Bradley*	Female condition: *Bridget*
A1				
Met	81.8% (*n* = 18)	72.7% (*n* = 16)	81.8% (*n* = 18)	77.3% (*n* = 17)
Partly met	18.2% (*n* = 4)	4.5% (*n* = 1)	9.1% (*n* = 2)	9.1% (*n* = 2)
Not met	0.0% (*n* = 0)	22.7% (*n* = 5)	9.1% (*n* = 2)	13.6% (*n* = 3)
Confidence met *M(SD)*	69.9 (25.1)	64.3 (39.9)	68.8 (27.9)	71.2 (23.5)
Severity *M(SD)*	56.7 (22.3)	39.9 (20.5)	50.5 (25.3)	53.9 (17.0)
A2				
Met	77.3% (*n* = 17)	81.0% (*n* = 17)	61.9% (*n* = 13)	72.7% (*n* = 16)
Partly met	9.1% (*n* = 2)	9.5% (*n* = 2)	28.6% (*n* = 6)	9.1% (*n* = 2)
Not met	13.6% (*n* = 3)	9.5% (*n* = 2)	9.5% (*n* = 2)	18.2% (*n* = 4)
Confidence met *M(SD)*	61.1 (27.7)	67.5 (17.8)	62.5 (24.5)	65.4 (23.4)
Severity *M(SD)*	48.3 (22.0)	41.3 (17.1)	43.1 (22.5)	48.9 (20.1)
A3				
Met	90.9% (*n* = 20)	81.0% (*n* = 17)	77.3% (*n* = 17)	90.9% (*n* = 20)
Partly met	4.5% (*n* = 1)	14.3% (*n* = 3)	13.6% (*n* = 3)	4.5% (*n* = 1)
Not met	4.5% (*n* = 1)	4.8% (*n* = 1)	9.1% (*n* = 2)	4.5% (*n* = 1)
Confidence met *M(SD)*	70.2 (25.9)	68.4 (17.6)	65.1 (24.7)	71.8 (17.7)
Severity *M(SD)*	54.4 (20.9)	48.0 (17.6)	51.8 (23.5)	57.8 (14.5)
Criteria A severity				
Not enough met	13.6% (*n* = 3)	9.1% (*n* = 2)	23.8% (*n* = 5)	15.0% (*n* = 3)
Level 1	31.8% (*n* = 7)	45.5% (*n* = 10)	33.3% (*n* = 7)	50.0% (*n* = 10)
Level 2	50.0% (*n* = 11)	31.8% (*n* = 7)	42.9% (*n* = 9)	35.0% (*n* = 7)
Level 3	4.5% (*n* = 1)	4.5% (*n* = 1)	0.0% (*n* = 0)	0.0% (*n* = 0)
B1				
Met	22.7% (*n* = 5)	27.3% (*n* = 6)	54.5% (*n* = 12)	70.0% (*n* = 14)
Partly met	36.4% (*n* = 8)	40.9% (*n* = 9)	31.8% (*n* = 7)	25.0% (*n* = 5)
Not met	40.9% (*n* = 9)	31.8% (*n* = 7)	13.6% (*n* = 3)	5.0% (*n* = 1)
Confidence met *M(SD)*	46.4 (30.0)	50.8 (24.5)	56.6 (19.0)	66.0 (22.6)
Severity *M(SD)*	28.7 (20.6)	23.7 (17.8)	34.4 (19.1)	41.5 (17.9)

	Female presentation case study		Male presentation case study	
Criterion	Female condition: *Grace*	Male condition: *Gordon*	Male condition: *Bradley*	Female condition: *Bridget*
B2				
Met	77.3% (*n* = 17)	71.4% (*n* = 15)	66.7% (*n* = 14)	72.7% (*n* = 16)
Partly met	22.7% (*n* = 5)	23.8% (*n* = 5)	19.0% (*n* = 4)	18.2% (*n* = 4)
Not met	0.0% (*n* = 0)	4.8% (*n* = 1)	14.3% (*n* = 3)	9.1% (*n* = 2)
Confidence met *M(SD)*	66.8 (23.5)	67.1 (20.3)	59.1 (23.8)	66.0 (21.1)
Severity *M(SD)*	50.5 (20.7)	43.5 (20.7)	41.9 (23.4)	44.4 (19.3)
B3				
Met	81.8% (*n* = 18)	63.6% (*n* = 14)	59.1% (*n* = 13)	59.1% (*n* = 13)
Partly met	4.5% (*n* = 1)	31.8% (*n* = 7)	13.6% (*n* = 3)	27.3% (*n* = 6)
Not met	13.6% (*n* = 3)	4.5% (*n* = 1)	27.3% (*n* = 6)	13.6% (*n* = 3)
Confidence met *M(SD)*	65.5 (28.5)	63.4 (20.8)	59.1 (25.1)	56.8 (25.6)
Severity *M(SD)*	45.7 (22.0)	34.9 (15.8)	42.3 (26.3)	44.9 (22.9)
B4				
Met	68.2% (*n* = 15)	66.7% (*n* = 14)	54.5% (*n* = 12)	72.7% (*n* = 11)
Partly met	18.2% (*n* = 4)	28.6% (*n* = 6)	36.4% (*n* = 8)	18.2% (*n* = 7)
Not met	13.6% (*n* = 3)	4.8% (*n* = 1)	9.1% (*n* = 2)	9.1% (*n* = 4)
Confidence met *M(SD)*	61.1 (27.4)	68.6 (13.4)	62.1 (20.5)	54.9 (26.6)
Severity *M(SD)*	40.3 (22.5)	39.2 (20.5)	38.8 (18.6)	34.8 (17.5)
Criteria B severity				
Not enough met	18.2% (*n* = 4)	13.6% (*n* = 3)	28.6% (*n* = 6)	20.0% (*n* = 4)
Level 1	45.5% (*n* = 10)	50.0% (*n* = 11)	38.1% (*n* = 8)	50.0% (*n* = 10)
Level 2	36.4% (*n* = 8)	27.3% (*n* = 6)	33.3% (*n* = 7)	30.0% (*n* = 6)
Level 3	0.0% (*n* = 0)	0.0% (*n* = 0)	0.0% (*n* = 0)	0.0% (*n* = 0)
C				
Met	95.5% (*n* = 21)	95.0% (*n* = 19)	85.7% (*n* = 18)	90.0% (*n* = 18)
Partly met	4.5% (*n* = 1)	0.0% (*n* = 0)	4.8% (*n* = 1)	10.0% (*n* = 2)
Not met	0% (*n* = 0)	5.0% (*n* = 1)	9.5% (*n* = 2)	0.0% (*n* = 0)
Confidence met *M(SD)*	73.0 (20.0)	66.7 (19.9)	69.0 (25.8)	71.9 (23.3)

	Female presentation case study		Male presentation case study	
Criterion	Female condition: *Grace*	Male condition: *Gordon*	Male condition: *Bradley*	Female condition: *Bridget*
D				
Met	77.3% (*n* = 17)	75.0% (*n* = 15)	71.4% (*n* = 15)	90.0% (*n* = 18)
Partly met	22.7% (*n* = 5)	10.0% (*n* = 2)	19.0% (*n* = 4)	5.0% (*n* = 1)
Not met	0.0% (*n* = 0)	15.0% (*n* = 3)	9.5% (*n* = 2)	5.0% (*n* = 1)
Confidence met *M(SD)*	71.9 (19.1)	65.8 (16.5)	65.5 (19.7)	70.9 (19.4)
E				
Met	81.8% (*n* = 18)	70.0% (*n* = 14)	81.0% (*n* = 17)	80.0% (*n* = 16)
Partly met	4.5% (*n* = 1)	5.0% (*n* = 1)	0.0% (*n* = 0)	5.0% (*n* = 1)
Not met	13.6% (*n* = 3)	25.0% (*n* = 5)	19.0% (*n* = 4)	15.0% (*n* = 3)
Confidence met *M(SD)*	66.0 (27.8)	66.7 (28.5)	72.9 (25.1)	66.5 (26.4)
ASD				
Yes	90.9% (*n* = 20)	85% (*n* = 17)	76.2% (*n* = 16)	80.0% (*n* = 16)
No	9.1% (*n* = 2)	15% (*n* = 3)	23.8% (*n* = 5)	20.0% (*n* = 4)
Confidence met *M(SD)*	66.0 (27.2)	66.0 (21.9)	65.6 (27.5)	62.2 (22.5)
Difficulty *M(SD)*	47.9 (25.9)	52.6 (23.9)	46.7 (19.7)	54.9 (18.7)

Note. Confidence represents diagnosticians' degree of certainty that a given criterion is met from 0 (*not at all confident*) to 100 (*extremely confident*). Severity indicates diagnosticians' impression of the degree of impairment in each domain from 0 (*very mild*) to 100 (*very severe*).

Table L.2

Frequency of Selecting Each Diagnosis and Differential/Additional Diagnosis

	Female presentation case study		Male presentation case study	
Diagnosis	Female condition *Grace* (*n* = 21)	Male condition *Gordon* (*n* = 20)	Male condition *Bradley* (*n* = 21)	Female condition *Bridget* (*n* = 18)
Diagnoses				
ASD	81.0% (*n* = 17)	80.0% (*n* = 16)	71.4% (*n* = 15)	77.8% (*n* = 14)
ADHD	0.0% (*n* = 0)	15.0% (*n* = 3)	33.3% (*n* = 7)	16.7% (*n* = 3)
GAD	19.0% (*n* = 4)	10.0% (*n* = 2)	4.8% (*n* = 1)	5.6% (*n* = 1)
Dev delay	4.8% (*n* = 1)	5.0% (*n* = 1)	0.0% (*n* = 0)	5.6% (*n* = 1)
ID	4.8% (*n* = 1)	0.0% (*n* = 0)	-	-
No Dx.	0.0% (*n* = 0)	15.0% (*n* = 3)	19.0% (*n* = 4)	16.7% (*n* = 3)
Additional/differential diagnoses				
ADHD	42.9% (*n* = 9)	50.0% (*n* = 10)	85.7% (*n* = 18)	61.1% (*n* = 11)
GAD	47.6% (*n* = 10)	65.0% (*n* = 13)	23.8% (*n* = 5)	5.6% (*n* = 1)
SocAD	52.4% (*n* = 11)	30.0% (*n* = 6)	9.5% (*n* = 2)	11.1% (*n* = 2)
OCD	9.5% (*n* = 2)	10.0% (*n* = 2)	4.8% (*n* = 1)	11.1% (*n* = 2)
SCD	33.3% (*n* = 7)	40.0% (*n* = 8)	47.6% (*n* = 10)	44.4% (*n* = 8)
LD	23.8% (*n* = 5)	35.0% (*n* = 7)	23.8% (*n* = 4)	22.2% (*n* = 4)
ID	28.6% (*n* = 6)	25.0% (*n* = 5)	14.3% (*n* = 3)	22.2% (*n* = 4)
SepAD	14.3% (*n* = 3)	15.0% (*n* = 3)	-	-
RAD	4.8% (*n* = 1)	10.0% (*n* = 2)	0.0% (*n* = 0)	11.1% (*n* = 2)
PTSD	9.5% (*n* = 2)	0.0% (*n* = 0)	4.8% (*n* = 1)	5.6% (*n* = 1)
CD	-	-	14.3% (*n* = 3)	5.6% (*n* = 1)
ODD	-	-	14.3% (*n* = 3)	11.1% (*n* = 2)
None	9.5% (*n* = 2)	10.0% (*n* = 2)	9.5% (*n* = 2)	22.2% (*n* = 4)
SLD*	4.8% (*n* = 1)	0.0% (*n* = 0)	-	-
SPD*	0.0% (*n* = 0)	5.0% (*n* = 1)	-	-
Dev delay*	4.8% (*n* = 1)	0.0% (*n* = 0)	-	-
FASD*	-	-	4.8% (*n* = 1)	0.0% (*n* = 0)

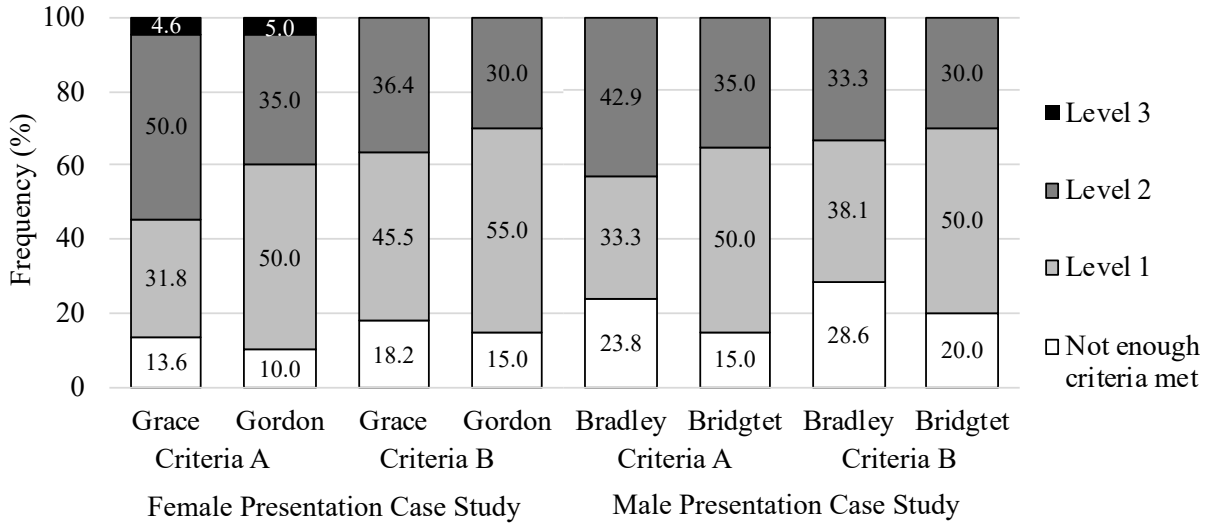
Note. Underlined values indicate which sex/gender condition was meaningfully more likely to receive the diagnosis according to the results of logistic regression.

Many diagnosticians suggested several diagnoses. Therefore, percentages do not sum to 100%.

* Diagnosis not supplied as an option, entered by diagnosticians as 'other (specified)'.
ADHD = attention-deficit/hyperactivity disorder, GAD = generalised anxiety disorder, SocAD = social anxiety disorder, OCD = obsessive-compulsive disorder, SCD = social (pragmatic) communication disorder, LD = language disorder, SepAD = separation anxiety disorder, RAD = reactive attachment disorder, PTSD = post-traumatic stress disorder, ID = intellectual disability, CD = conduct disorder, ODD = oppositional defiance disorder, BPD = borderline personality disorder, SLD = specific learning disorder, SPD = sensory processing disorder, Dev delay = developmental delay, FASD = foetal alcohol spectrum disorder.

Figure L

Frequency of Support Level Selections in Criteria A and B for Each Case Study and Condition



Note. This figure is derived from raw proportions data (rather than the model) and therefore HDIs $_{(80\%)}$ (etc.) are not included. Criteria A: Social communication; Criteria B: Repetitive and restricted behaviours and interests.

Table L.3

Results of Logistic Regressions: Diagnoses and Differential/Additional Diagnoses

	Intercept	Case Study	Sex/Gender Condition	Interaction
	LOR [$HDI_{80\%}$]			
Diagnosis				
ASD	1.38 [1.01, 1.73]	-0.31 [-1.03, 0.40]	-0.31 [-1.05, 0.39]	-0.38 [-1.79, 1.02]
ADHD	-2.68 [-3.42, -1.80]	2.71 [0.98, 4.19]	2.70 [-.95, 4.11]	-3.01 [-5.85, 0.32]
GAD	-2.72 [-3.27, -1.80]	-1.40 [-2.53, -0.10]	-0.44 [-1.60, 0.81]	0.70 [-1.57, 3.12]
Dev. delay	-4.34 [-5.34, -3.09]	-1.59 [-3.64, 0.75]	-1.56 [-3.54, 0.76]	-2.81 [-6.33, 1.29]
ID	-7.32 [-9.41, -4.66]	-2.95 [-6.79, 1.14]	-2.76 [-6.35, 1.41]	2.38 [-3.00, 7.62]
No Dx.	-2.91 [-3.64, -1.95]	2.35 [0.47, 3.75]	2.35 [0.45, 3.75]	-3.86 [-6.74, -0.37]
Additional/differential diagnoses				
ADHD	0.52 [0.20, 0.86]	1.33 [0.66, 1.99]	0.87 [0.19, 1.52]	1.10 [-0.28, 2.32]
GAD	-0.96 [-1.36, -0.49]	-2.45 [-3.26, -1.52]	1.32 [0.39, 2.12]	1.12 [-0.64, 2.74]
SocAD	-1.38 [-1.80, -0.95]	-1.96 [-2.77, -1.09]	-0.59 [-1.45, -0.22]	0.79 [-0.83, 2.45]
OCD	-2.66 [-3.21, -2.04]	-0.40 [-1.54, 0.76]	-0.53 [-1.68, 0.64]	-1.08 [-3.21, 1.19]
SCD	-0.37 [-0.68, -0.08]	0.41 [-0.19, 1.01]	0.22 [-0.37, 0.83]	-0.16 [-1.36, 1.01]
LD	-1.19 [-1.54, -0.84]	-0.49 [-1.19, 0.20]	0.20 [-0.50, 0.90]	-0.74 [-2.10, 0.64]
ID	-1.35 [-1.69, -0.96]	-0.56 [-1.27, 0.18]	-0.39 [-1.11, 0.35]	-0.36 [-1.77, 1.09]

Note. The intercept column indicates the frequency of diagnosis endorsement (higher numbers = greater frequency). The case study column conveys the role the case study (female or male; positive = male presentation) and the sex/gender condition column shows the effect of the condition (positive = male sex/gender condition) on the probability of diagnosis endorsement. Finally, the interaction column shows the case study × sex/gender condition interaction.

LOR = log odds ratio, ADHD = attention-deficit/hyperactivity disorder, GAD = generalised anxiety disorder, SocAD = social anxiety disorder, OCD = obsessive-compulsive disorder, SCD = social (pragmatic) communication disorder, LD = language disorder, ID = intellectual disability.

Appendix M: Categories for Responses to Two Open-Ended Questions (Study 3)

Table M.1

Reasons ASD may be Underdiagnosed in Females (n = 40 diagnosticians, 115 responses)

Category	Responses *(n)*	Example Quotation
1. Sex/gender differences in ASD presentation:	75	
(a) Camouflaging: masking and compensation	35	"Girls may be better at masking their difficulties e.g., using imitation."
(b) More subtle/less obvious impairment	10	"[Females may present with] subtle repetitive behaviours."
(c) More socially appropriate special interests	10	"[Females may have] more socially acceptable special interests (e.g., animals, music)."
(d) Less disruptive/ externalising behaviour	9	"Females in general may show less externalising behaviour- so their challenges go unnoticed."
(e) More socially motivated	7	"Increased interest in peers/desire for friendships (just lack skills to develop and maintain them)."
(f) Different expression of symptoms	4	"Differences in presentation at assessment."
2. The female presentation of ASD remains poorly understood and under-researched	10	"Lack of understanding of their presentation by professionals."
3. Bias in ASD assessment tools and conceptualisation	8	"Diagnostic tools miss the female autism phenotype."
4. Professionals do not look for ASD in females: gender expectations and priming	7	"Professionals have pre-conceptions of [the] autism 'presentation'."
5. Misdiagnosis or diagnostic overshadowing	6	"Clinicians can struggle to differentiate between anxiety and ASD in young girls…"
6. Girls' presentations may vary across different environments and with time	5	"Data across environments is often at odds… especially when girls are able to 'hold onto' their reactions at school. They typically present when social environment exceeds capacity, which I find is often 8+ years, thus diagnosed later than [the] average male."

Category	Responses *(n)*	Example Quotation
7. Normative sex/gender differences exist in neurology and developmental trajectories	4	"[Girls] may have a developmental advantage in social engagement."

Note. Not all diagnosticians responded to this question and some provided fewer than three responses.

Table M.2

Categories Identified for Greatest Challenges in Assessing Females for ASD (n = 38, 66 Responses)

Category	Responses *(n)*	Example Quotation
1. Negotiating the mismatch in symptom manifestation with 'classic' ASD:	23	
(a) Assessment tools/ASD conceptualisation	16	"I lack trust in scores on instruments and tools when working with females, knowing they've been developed with a male bias… I'm not sure there exists a clear conceptualisation of autism in females, meaning some girls might be missed."
(b) More subtle presentation, better social skills	5	"Identifying idiosyncratic and subtle manifestations of criteria."
(c) Qualitative differences in obsessive interests	2	"Their interests may be more socially acceptable… whereas boys' [interests] are less mainstream."
2. Recognising camouflaging	20	"It's challenging to know if [girls] have just learnt very good skills in acting like others or if this has naturally developed."
3. Establishing a differential diagnosis	9	"Often female autistics [sic] seem to be misdiagnosed with social anxiety, borderline personality disorder or another psychiatric illness. Teasing apart these conditions can be challenging, especially since almost all female autistics have comorbid anxiety and/or depression as a result of their social difficulties."
4. Limited knowledge of, and differences in professionals' opinions about the female ASD presentation	9	"When talking to parents and teachers you need to understand ASD in girls to be able to get the accurate information to lead to an accurate assessment/result."
5. Reconciling differences in presentation across settings	6	"Schoolteachers often do not report any concerns regarding child functioning within school environment."

Note. Not all diagnosticians responded to this question and some provided fewer than three responses.