Understanding executive function in young autistic people: moving from the lab to the everyday

Lorcan Kenny

UCL Institute of Education

Thesis submitted for the degree of Doctor of Philosophy (PhD) in Psychology in September 2018

Abstract

A wealth of empirical research has focused on the executive function (EF) ability of autistic people. This research initially focused predominantly on establishing whether EF difficulties could be causal to the development of autistic traits. More recently, the focus has been on establishing the role of EF in explaining the life chances of autistic people. In this thesis, I aimed to test whether the conclusions drawn from traditional laboratory-based EF assessments have a bearing on the everyday lives of autistic young people. To address this aim, In Chapter 2, I examined whether the group differences between autistic and non-autistic participants on EF assessments commonly reported in the literature result from differences in the latent executive constructs underlying task performance or from task-specific, non-executive characteristics. In Chapter 3, I assessed whether group differences in EF ability emerged on a representatively-designed measure of EF. In Chapter 4, I investigated whether other metrics of real-world functional adjustment were related to individual differences in performance on both the systematically-designed and on the representatively-designed EF tasks, described in Chapters 3 and 4, respectively. In Chapter 5, I sought to understand the personal and parental perspectives about the executive abilities of autistic young people and the realities of these (dis)abilities on everyday life. Finally, in Chapter 6, I discuss how the findings from this thesis relate to the extant literature before making five recommendations on approaches I believe will move the discipline of EF research with autistic people forward. Namely, that to understand how EF difficulties relate to the lives of autistic people, future research should; (i) validate systematically-designed tasks against real-world measures with autistic populations, (ii) employ representatively-designed tasks to capture functional abilities, (iii) triangulate objective assessment with subjective reports, (iv) conduct carefully controlled experiments to elucidate the mechanisms underlying everyday executive difficulties and

(v) conceptualise of EF in autism within a dimensional, rather than a categorical, framework.

Impact Statement

There are four ways in which the work presented within this thesis could generate impact. First, the analyses and discoveries presented within this thesis have the potential to reshape future scholarship in the field of executive function research with autistic people. The arguments presented within this thesis, alongside the results described, set forth an approach to conducting research that has a greater likelihood of discovering ways in which the quality of life of autistic people who struggle with executive function abilities can be improved upon. These findings have been presented to leading autism researchers at the 2018 Annual Meeting of the International Society for Autism Research in Rotterdam, Netherlands.

Second, the instructions for administering and scoring performance on the Modified Multitasking Evaluation for Adolescents, a significant modification of an existing assessment that has been presented in this thesis, have been deposited in a publicly accessible repository. This is to facilitate direct replication attempts and future investigations by other researchers. Similarly, the NIH-EXAMINER battery described in this thesis was selected because it is open source and, consequently, direct replication attempts are possible. Third, de-identified data collected and analysed within this thesis have been made available in the same public repository. Making data openly available will facilitate re-analysis of the data, allow more reliable synthesis of evidence and make cumulative data collection with iterative additions by other researchers, increasing the accuracy of conclusions and minimising the duplication of effort.

Fourth, the research conducted for the purposes of this thesis have had an impact on public discourse. Data were collected during a public engagement event where children and young people learned about psychological research. Finally, an adapted version of the naturalistic executive function assessment research reported on in Chapter 3 and in Chapter 4 featured in a documentary called 'Are you Autistic?'

which was broadcast on primetime national television to an audience in excess of 1.6 million viewers.

Declaration

I, Lorcan Kenny, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. The Multitasking Evaluation for Adolescents described in this thesis was developed by Dr Maude Schneider and adapted for use in the studies presented within this thesis. Much of the data presented here were collected at Brain Detectives public engagement and science participation events at UCL Institute of Education and the Brain Detectives Team assisted with scheduling participants, administering IQ assessments and data input. Sarah Crockford and Dr Melissa Bovis occasionally provided assistance with the collection of data on the NIH-EXAMINER battery and Zora Lattoe conducted two of the qualitative interviews included.

Word count (excluding references): 66,919 words

Signature: ______________________________

Date: ______________________________

Acknowledgements

This thesis would not have been possible without the autistic and non-autistic young people and their parents whose untiring commitment to my research was truly humbling. It was a privilege to work with all of those who gave so freely of their time and who shared so unreservedly of their experiences. The work in this thesis was generously supported by an Economic and Social Research Council PhD studentship and the work at the Centre for Research in Autism and Education (CRAE) is supported by the Pears Foundation.

I would like to thank my supervisors Liz Pellicano and Anna Remington for their guidance, support and ceaseless generosity with their time throughout this process. I am grateful to have worked with such an amazing team at CRAE, in particular, to all of those involved in organising and running the Brain Detectives workshops, where much of the data presented in this thesis were collected. In particular, I thank Sarah Crockford, Mel Bovis, Eleanor Palser, Eilish Roy, Rachel Prosser and Demi Eades.

Above all, I owe a debt of gratitude to my family and friends who supported me throughout my PhD. I owe thanks to my siblings, Saoirse and Eoghan, who taught me the art of argument around the dinner table, to my climbing buddies, Liz and Sam, who gave me an outlet to unwind during my studies, to all those with whom I shared a coffee, trip to the farmers' market or a supportive chat, including, Shane, Mags, Lorna, Mel, Ellie, Jana and James. I am grateful to my parents, Geraldine and Conor, who instilled in me the value of education, who cultivated in me a natural curiosity, who gave me the work ethic to pursue that curiosity and who quietly sacrificed so much so that no opportunity was ever out of my reach. Finally, I owe thanks that is above and beyond anything I can articulate to my girlfriend, Aoife, who believed in me when my belief in myself wavered, whose words of encouragement kept me writing, who listened to me when I needed it and who knew to ignore me when I needed that.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
ADOS-2	Autism Diagnostic Observation Schedule, second edition
ADI-R	Autism Diagnostic Interview - Revised
ADS	Action Disorganisation Syndrome
APA	American Psychiatric Association
BADS	Behavioural Assessment of the Dysexecutive Syndrome
BADS-C	Behavioural Assessment of the Dysexecutive Syndrome for children
BRIEF	Behavior Rating Inventory of Executive Function
BRIEF-A	Behavior Rating Inventory of Executive Function, adult version
BRIEF-2	Behavior Rating Inventory of Executive Function, second edition
CANTAB	Cambridge Neuropsychological Test Automated Battery
CDC	Centers for Disease Control and Prevention
CPT	Continuous performance task
CRAE	Centre for Research in Autism and Education
DCCS	Dimensional Change Card Sort task
DES	Dysexecutive syndrome
DEX	Dysexecutive syndrome questionnaire
D-KEFS	The Delis–Kaplan Executive Function System Design Fluency Test
DSM-III	Diagnostic and Statistical Manual of Mental Disorders, third edition
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, fourth edition
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, fifth edition
d`	D-prime
EF	Executive function
FrsBe	Frontal Systems Behaviour Scale
NEPSY-II	Developmental Neuropsychological Assessment
NIH	National Institute of Health
EXAMINER	Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research
NIMH	National Institute of Mental Health
ICD-10	The International Classification of Mental and Behavioural Disorders, tenth edition
ICD-11	The International Classification of Mental and Behavioural Disorders, eleventh edition
ICF	The International Classification of Functioning, Disability and Health
ID	Intellectual disability
ID/ED	Intra-dimensional/extra-dimensional shifting task
IMD	Index of Multiple Deprivation
OCD	Obsessive compulsive disorder
ODD	Oppositional defiant disorder
PDD-NOS	Pervasive Developmental Disorder Not Otherwise Specified
QoL	Quality of life
RDoC	Research Domain Criteria
RRBs	Restricted, repetitive patterns of behaviour or interests
RT	Reaction time
SES	Socio-economic status
SRS-2	Social Responsiveness Scale, second edition
ToM	Theory of mind
WCST	Wisconsin Card Sorting Task
WHO	The World Health Organization
WHODAS-2	The World Health Organization Disability Assessment Schedule, second edition

Chapter 1

General Introduction

This thesis focuses on whether a set of cognitive abilities that include working memory, flexibility and inhibition, collectively referred to as executive function (EF), differ for autistic[1] people. Specifically, it will test whether a measure of EF designed to be representative of a real-world environment in which EF might typically be employed is especially sensitive to group differences between autistic and non-autistic adolescents. Furthermore, it will consider whether individual differences in EF ability in a task that represents real-world executive demands or in more traditional neuropsychological EF tasks are differentially related to variation in quality of life, mental health or self-reported (dis)ability. Finally, it will provide an opportunity for young autistic people to share their own experiences of their EF abilities and to add their voices to a literature that has hitherto excluded their perspectives.

In the present chapter, I will consider what autism is, how it is diagnosed, and what we know about its causes and brain basis. I will then detail cognitive theories of autism, which serve to link the biological and behavioural profiles in autism. Finally, I review in detail the literature that focuses on what EF is, how it is measured and its relationship with autism before setting out the rationale for the studies that form this thesis.

[1] In this thesis, I use identity-first language (e.g. 'autistic person') rather than person-first language (e.g., 'person with autism', 'person who has autism') in deference to the majority of autistic people (Kenny et al., 2016; Sinclair, 1999) and to reduce stigma (Gernsbacher, 2017).

What is autism?

History of autism. In approximately 1908 a Swiss Psychiatrist named Eugen Bleuler coined the term 'autism' in his Zürich clinic. The first published record of autism was when Bleuler (1911) referred to a specific cluster of behaviours that a subset of patients with schizophrenia showed, 'autistic detachment', which, according to Bleuler was a vital "detachment from reality with the relative and absolute predominance of the inner life" (p. 63). It was not until the 1940s, though, when autistic behaviour began to be considered a distinct condition worthy of consideration. In a now seminal paper, Leo Kanner (1943) presented eleven case studies of children who all appeared to show some degree of detachment from the social world, overly routinised or ritualistic behaviour and, in some cases, an ambivalence towards communicative cues that would typically command attention.

Within months of this paper's publication, a Viennese Psychiatrist named Hans Asperger (1944) published a paper entitled 'Autistic psychopathy in childhood' with some striking similarities to Kanner's case descriptions, including that they both independently used the term 'autism'. There were, however, some important discrepancies in these early accounts. Asperger believed autism was not a rare condition and that it could affect people on a continuum of cognitive ability, whereas Kanner's definition was much narrower and many of his case studies focused on children who had little or no expressive language. Further, Asperger believed autism had a neurological cause while Kanner's psycho-analytic contemporaries inspired him to suggest that maternal coldness was the trigger for the behaviour he had been observing (Kanner, 1949) – a suggestion that paved the way for a now discredited but nonetheless continually damaging *Refrigerator Mother Hypothesis* (Bettelheim, 1959).

These early descriptions of autism were the bedrock which ultimately led to the inclusion of *infantile autism* in the Diagnostic and Statistical Manual for Mental Disorders, third edition (DSM-III; American Psychiatric Association [APA], 1980), for the first time, facilitating autism diagnoses to be made. Lorna Wing (Wing, 1981) described autism as a spectrum condition incorporating both Kanner and Asperger's descriptions whereby people had a constellation of symptoms within three domains: social interaction, social imagination and communication. Accordingly, the fourth edition of the manual (DSM-IV; APA, 1994), replaced infantile autism with a spectrum of related neurodevelopmental conditions, under the umbrella term 'pervasive developmental disorder', including autistic disorder, Asperger's disorder and pervasive developmental disorder not otherwise specified (PDD-NOS).

The diagnosis of autism relied on a person having behavioural difficulties beginning in childhood including marked difficulties with interaction, verbal or non-verbal communication and restricted, repetitive and stereotyped behaviours or interests (APA, 2000). The distinction between Asperger's disorder and autistic disorder, the latter of which more closely matched Kanner's (1943) description, was drawn based on the absence of an intellectual disability (ID) or language delay in Asperger's disorder. A PDD-NOS diagnosis, instead, was given when some but not all the criteria for autistic disorder or Asperger's disorder were met. The DSM-IV criteria were, however, replaced because research repeatedly failed to reliably distinguish between these supposed subtypes of autism (Grzadzinski, Huerta, & Lord, 2013) and diagnostic practises were inconsistently applied across clinical settings (Lord, Petkova, Hus, Gan, & Lu, 2012).

Current diagnostic features of autism. The 2013 publication of the fifth, revised edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013) brought considerable changes to the diagnostic category[2], including renaming Pervasive Developmental Disorders as Autism Spectrum Disorder (hereafter, 'autism'). Autism, as defined in DSM-5, involves persistent difficulties in two, rather

---

[2] There is an alternative diagnostic manual published by the World Health Organization, called the International Classification of Diseases, eleventh edition (ICD-11; WHO, 2018). The ICD-11 largely mirrors the DSM-5 except that it provides more detailed instructions to differentiate between those with and without an intellectual disability and places less emphasis on the type of play engaged in by children, to reduce the influence of cultural differences on diagnosis.

than three, domains: 'social communication and interaction' and 'restricted, repetitive patterns of behaviours or interests' (RRBs; APA, 2013). The first of these, often referred to as the social domain, describes an atypical pattern of social behaviour involving, each of the following three types of social difficulties. First, having difficulties with social reciprocity, for example, reduced inclination toward social engagement, unusually one-sided conversations or interactions that are not well sustained. Second, people on the autism spectrum are also less likely to naturally integrate eye gaze and gestures with speech during social interactions. Third, autistic people can also have difficulties understanding the nature of typical social relationships, for instance, the differences between a friend and romantic partner, and problems developing and sustaining such relationships.

The second, so-called non-social, domain of autism symptoms describes specific behaviour that must be present, in addition to that described above, to receive an autism diagnosis. The behaviour described in this domain includes: (i) stereotyped or repetitive motor movements; (ii) an insistence on sameness or inflexible adherence to routines; (iii) highly restricted, fixated interests and (iv) hyper- or hypo-reactivity to sensory input or an unusual interest in sensory aspects of the environment. At least two out of these four types of behaviours must be present for somebody to receiver a diagnosis. These difficulties could be apparent, for example, when an autistic person systematically lines up items (e.g., toy cars), engages in verbal routines like compulsively listing items (e.g., names of stations on the London Underground network) or shows an intense interest in a pattern (e.g., the shapes created by a spinning top). Others on the autism spectrum might display these non-social features through an inflexible observance of a routine, for example, eating the same food, at the same time, in the same place each day. Another person, instead, might be compelled to obey ritualised behaviours, for example, being unable to leave a classroom until all the chairs are neatly under the tables. Unexpected changes to routines or the inability to complete a ritual can lead to distress, or in some cases can precipitate a meltdown (an intense reaction to an overwhelming situation) or shutdown (a retreat inward to cope with an overwhelming situation). Showing restricted, intense interests, for example, an affinity for Disney sidekicks or a highly-specific, unusual interest in the inner workings of door locks, is another characteristic behaviour of autism. The DSM-5 introduced unusual sensory experiences as a core diagnostic symptom of autism for the first time, reflecting the considerable impact sensory differences have on the lives of autistic people[3].

Autism is a developmental rather than an acquired condition, and core symptoms must therefore be present in the first years of life (APA, 2013). Across development it becomes increasingly likely that learned skills mask symptoms, complicating assessment and diagnosis in adolescence or adulthood (Livingston & Happé, 2017). The absence of a reliable biological diagnostic assessment means there is no objective threshold where a neat distinction is made between being autistic and being non-autistic. Instead, the distinction is based on whether a person has a constellation of the social and non-social behavioural symptoms described above and, crucially, that these symptoms cause 'clinically significant' impairments in social and/or occupational functioning; a decision that is typically reached by a multi-disciplinary team of clinicians.

The specific cluster of autistic features that combine to create a clinically significant impairment varies widely from person to person, resulting in heterogeneous symptom profiles (Mandell, 2011). To help demarcate this heterogeneity, clinicians specify a person's symptom severity in each domain based on whether they are 'requiring support', 'requiring substantial support', or 'requiring very substantial support' (APA, 2013). Clinicians use observation tools such as the Autism Diagnostic Observation Schedule – Second edition (ADOS-2; Lord, Rutter, et al., 2012) and

[3] While DSM-5 included sensory sensitivity as a core diagnostic autism trait, the ICD-11 did not include sensory sensitivity and so diagnosis may vary as a function of geographical location and diagnostic manual employed.

interview tools such as the Autism Diagnostic Interview – Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) to help in the decision-making process surrounding autism diagnosis. Observation and interview tools are often supplemented with questionnaire measures of autism symptomatology such as the Social Responsiveness Scale – Second edition (SRS-2; Constantino & Gruber, 2012), the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) or the Autism Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001).

Co-occurring conditions. As well as heterogeneous profiles of core autism features, autism very often co-occurs with a range of other conditions; something referred to in the medical literature as 'comorbidity' (Leyfer et al., 2006). A DSM-5 autism diagnosis should specify whether somebody has any accompanying language, intellectual, neurodevelopmental, mental, genetic or behavioural conditions (APA, 2013). Here, I introduce the most common conditions that co-occur with autism.

Co-occurring intellectual disabilities. The likelihood of an autistic person having an intellectual disability (ID) is much higher than that of a non-autistic person, with between 30 and 50% of autistic people estimated to have an ID (Centers for Disease Control and Prevention [CDC], 2014; Charman et al., 2011; Idring et al., 2012). Autism does not, however, have a specific profile of intellectual ability; autistic people span the whole range of possible intellects from profound ID to exceptional ability (Charman et al., 2011). ID has an onset during the developmental period that includes both intellectual *and* adaptive functioning difficulties in conceptual, social and practical domains (APA, 2013). IDs are defined as a score that falls two standard deviations or more below the population mean on a standardised IQ test, in other words, obtaining an IQ score of 70 or below. Adaptive functioning difficulties are defined as a failure to meet developmental and sociocultural standards for personal independence and social responsibility (APA, 2013). My interest is to understand whether autistic young people who do not have an ID are more likely than non-autistic young people to have specific, subtle cognitive difficulties, namely with EF, that can impact on their daily lives. I am interested in these skills because, unlike IQ, they are rarely assessed in clinical practice. The consequence is that if these skills do impact on people's life chances but often go undetected then supports that may mitigate against these difficulties are not often tested in the scientific literature or applied in daily life. This thesis therefore focuses on those who do not meet the criteria for an ID, based on having an IQ within at least the average range (i.e., IQ scores greater than 70), although these individuals may nevertheless have difficulties with adaptive functioning. I will refer to this group henceforth as being 'cognitively able'.

Co-occurring psychiatric diagnoses. People on the autism spectrum also have an increased likelihood of experiencing co-occurring psychiatric conditions. One UK population-based study found that 70% of autistic children had at least one such co-occurring condition, 40% had at least two (Simonoff et al., 2008), the most common of which were anxiety disorders (42%), oppositional defiant disorder (ODD; 30%) and attention-deficit hyperactivity disorder (ADHD; 28%). Rates of co-occurring conditions are likely substantially higher than these figures when samples are not population-based but instead recruited in the clinic or community. For example, when a sample of autistic children accessing a psychopharmacology service were systematically screened for co-occurring conditions, 95% of them had three or more additional psychiatric conditions, 74% had five or more (Joshi et al., 2010). Neither cognitive ability nor autism symptom severity is sufficient to predict comorbidity rates (Simonoff et al., 2008; Strang et al., 2012)

The high rate of comorbidity in autism persists well into adolescence and adulthood (Joshi et al., 2013; Simonoff et al., 2013). The most commonly reported co-occurring conditions were mood and anxiety disorders, both of which are reported to affect approximately half of autistic adults at any given time (Joshi et al., 2013; Lever & Geurts, 2016). When lifetime occurrence is assessed this number rises to approximately

70% (Lugnegård, Hallerbäck, & Gillberg, 2011). This rate is markedly high compared with typically developing individuals (Lever & Geurts, 2016) and is twice as high, on average, as the rate among psychiatrically referred, non-autistic adults (Joshi et al., 2013). The reasons why autistic people have higher rates of other conditions is not well understood, a recent suggestion has been that we may gain more understanding of these co-occurrences by using temporal language to describe whether a condition is 'truly co-occurring', 'resulting from', or 'associated with' autism (Rubenstein & Bishop-Fitzpatrick, 2018). One fundamental, although hitherto unanswered, question remains: do these patterns of psychiatric symptoms in autism represent phenotypic manifestations of autism or are they instead an expression of distinct, but co-occurring conditions (Mazzone, Ruta, & Reale, 2012)?

Diagnosing co-occurring psychiatric conditions in autism remains a challenge, as diagnostic criteria rarely account for the autism-specific manifestations of these conditions. When a diagnostic tool is adapted to reduce the likelihood of misidentifying autistic features as symptoms from other conditions, for example, ODD, the criteria for additional diagnoses are often no longer met (Mazefsky et al., 2012). Further, many autistic people report finding it difficult to determine if they are experiencing mental health difficulties or if they are experiencing difficulties associated with being autistic in a world designed for non-autistic people (Crane, Adams, Harper, Welch, & Pellicano, 2018). Recognising and supporting these additional difficulties is critical. For many autistic people and their allies, their mental health difficulties, the associated stigma and a lack of public accommodations can have a greater perceived impact on their lives than does their being autistic (Ozsivadjian, Knott, & Magiati, 2012).

Medical comorbidities. Young people on the autism spectrum have been shown to have higher rates of many medical conditions including seizure, sleep disorders, gastrointestinal and metabolic disorders, hormone dysfunction, obesity and nutritional deficits, when compared with non-autistic children and adolescents (Bauman, 2010). Similarly, a comparison of 1,500 autistic and 15,000 non-autistic people demonstrated that almost all medical conditions were more common in the autistic population (Croen et al., 2015). Croen et al.'s comparison of privately insured adults in the US showed that among those on the autism spectrum, the following conditions occur more often: immune, gastrointestinal, sleep and blood disorders, seizure, obesity and diabetes. The higher incidence in autism remains for relatively rare conditions such as stroke and Parkinson's disease (see Mannion & Leader, 2013, for a review). Additionally, the presence or absence of co-occurring conditions appears to impact the age at which somebody is first evaluated for autism (Soke, Maenner, Christensen, Kurzius-Spencer, & Schieve, 2018). Consequently, research about autism must consider co-occurring intellectual, psychiatric and medical conditions.

Prevalence

Presently, between one and two per cent of the population are estimated to be autistic (Baird et al., 2006; Brugha et al., 2011; Idring et al., 2012; Kim et al., 2011; The Centers for Disease Control and Prevention, 2014). The prevalence of autism has rapidly increased over the past two decades. For example, the CDC in the United States have reported prevalence rates increasing from 1 in 150 in 2002 (The Centers for Disease Control and Prevention, 2007), to 1 in 110 in 2006 (The Centers for Disease Control and Prevention, 2009), to 1 in 88 in 2008 (The Centers for Disease Control and Prevention, 2012) with the latest estimate remaining stable at 1 in 68 between 2010 and 2012 (The Centers for Disease Control and Prevention, 2014, 2016). Unlike the United States, an analysis of the UK General Practice Research Database showed that prevalence rates remained stable in the UK between 2004 and 2010 (Taylor, Jick, & MacLaughlin, 2013). Whether there is a true rise in autism incidence is not fully understood but the evidence suggests that rising prevalence can mostly be accounted for by methodological issues in early studies, changes in diagnostic practices, increasing practitioner and public awareness of autism, and diagnostic substitution (Gernsbacher, Dawson, & Goldsmith, 2005; Hansen, Schendel, & Parner, 2015; Rice et al., 2012; Shattuck, 2006). This conclusion is bolstered by broadly similar prevalence rates worldwide, which appear to vary as a function of autism awareness, service availability and broadening diagnostic concepts (Elsabbagh et al., 2012).

Sex differences in prevalence rates. According to DSM-5, autism is "diagnosed four times more in males than in females" (APA, 2013, p. 57). This commonly-reported male-to-female ratio is found with striking regularity in epidemiological studies (see Fombonne, 2009, for review), although it is lower in samples of autistic individuals with co-occurring IDs (Skuse et al., 2009). A recent meta-analysis supported a male preponderance in diagnoses but showed that, in higher quality studies, particularly those employing active case ascertainment, male-to-female ratios were closer to 3:1 (Loomes, Hull, & Mandy, 2017). There is compelling evidence that the female sex confers a protective effect against autistic behaviour; females require a greater inherited etiological load before expressing the autism phenotype (Jacquemont et al., 2014; E. B. Robinson, Lichtenstein, Anckarsäter, Happé, & Ronald, 2013; Skuse, 2007). One potential biological candidate to explain the female protective effect is that sex hormones, testosterone in particular, may affect the liability for developing autism (Auyeung, Taylor, Hackett, & Baron-Cohen, 2010; Baron-Cohen, 2002).

Despite the likely role of biological explanations for the male-to-female ratio, the discrepancy may be exacerbated by longstanding, systematic diagnostic biases that render male-centric diagnostic instruments inappropriate for women and girls (Loomes et al., 2017; Mandy et al., 2012). Stereotypes about gender normative behaviour and about autism being a male condition (Bargiela, Steward, & Mandy, 2016) combined with autistic females' relatively greater sociality and increased likelihood of masking their difficulties, known as 'camouflaging' (Hull et al., 2017; Lai et al., 2016; Mandy et al., 2012), could result in people being less sensitive to autistic traits in females. Subtle differences in the behavioural manifestation of autism in cognitively able women and girls, for example, showing less externalising behaviour, mean women and girls may be less likely to be detected with standard diagnostic instruments (Hiller, Young, & Weber, 2014; Sedgewick, Hill, Yates, Pickering, & Pellicano, 2016; Werling & Geschwind, 2013). The cumulative effect of which has resulted in a generation of autistic females who were misdiagnosed, diagnosed at older ages or who missed out on a diagnosis all together (Werling & Geschwind, 2013).

Aetiology

There is a strong genetic liability in the aetiology of autism, as is evident from familial recurrence rates (Sandin et al., 2014) and twin studies (Bailey et al., 1995; Folstein & Rutter, 1977; Rosenberg et al., 2009). Alongside heritable genetic contributions, rare non-inherited genetic lesions, known as de-novo mutations, have been shown to be causally related to autism (Ronemus, Iossifov, Levy, & Wigler, 2014), albeit to a lesser extent than common heritable variants (Gaugler et al., 2014). Autism is predicted to involve more than 1,000 genes (De Rubeis et al., 2014), although certain monogenic disorders (e.g., Fragile X syndrome, Tuberous Sclerosis) are associated with autistic symptoms (Tsai & Sahin, 2011). The diversity of autism aetiologies have led to calls by some to consider autism as a grouping of conditions collectively known as "the autisms" and not as a monolithic condition (Geschwind & Levitt, 2007; London, 2014). Environmental factors such as increasing parental age, birth complications, caesarean section as well as obesity and diabetes during pregnancy also have a modest role in the aetiology of autism (for a review, see Modabbernia, Velthorst, & Reichenberg, 2017), which have complex relationships with genetic liability through a series of gene-environment interactions and gene-environment correlations (Mandy & Lai, 2016).

Dimensional psychological approaches. The DSM was introduced to bring about reliable diagnosis, but it has been repeatedly criticised for adopting an eminence-based rather than evidence-based approach (London, 2014). That is, the DSM describes observable individual characteristics (called phenotypes), sorts them into clusters (called symptoms) and people decide the distinct categories (called diagnoses) to which the symptoms belong. The decisions underlying this system are based on the consensus of a group of people deemed by the APA to be sufficiently eminent to make such decisions. This categorical classification system is useful insofar as it provides a shared nomenclature to clinicians and researchers but may be impeding scientific progress because, despite the reification of diagnoses, evidence for their reliability and validity remains wanting (Hyman, 2010). For instance, DSM categories tell us little about aetiology; lack reliable biological markers; have widely varying and poorly predictable prognoses; have multiple comorbidities, often resulting in diagnostic overshadowing; lack clear boundaries between disorder and typicality or between different disorders; symptoms lack rigorous definitions; and conditions are not predictably responsive to intervention (L. A. Clark, Watson, & Reynold, 1995; London, 2014; Mason & Scior, 2004).

The above-mentioned concerns also apply specifically to autism. First, DSM categorisation has resulted in limited understanding of the aetiology of autism. Second, the absence of reliable biological markers necessitates behavioural diagnosis (Volkmar et al., 2014). Third, prognosis is highly variable, and we do not know which factors predict it well. Fourth, as discussed above, autism has multiple comorbidities. One diagnosis often overshadows another, for example, an autistic person with excessive worry might be significantly more affected by their anxiety than by being autistic (Damiano, Mazefsky, White, & Dichter, 2014). Furthermore, we do not understand whether one developmental condition increases the likelihood of having another, whether misdiagnosis is common, or whether comorbidity is the result of diagnostic categories that do not reflect the true nature of these conditions (Caron & Rutter, 1991). Fifth, the distinction between being autistic and non-autistic is blurry, behavioural autistic traits are normally distributed in the population and the delimitation between autism, the 'broader autism phenotype' and non-autistic behaviour is arbitrary (Lundström et al., 2012). Finally, symptom descriptions remain vague and subjective. The clustering of symptoms necessary and sufficient for diagnosis is also imprecise and consequently people on the autism spectrum often show non-overlapping aetiologies and behavioural presentations (Betancur, 2011).

The International Classification of Functioning, Disability and Health. Due to the shortcomings listed above, diagnostic labels alone do not provide sufficient information about how a condition, such as autism, impacts on an individual's daily functioning, their self-perceived disability and by extension, their quality of life (Bölte et al., 2014). The International Classification of Functioning, Disability and Health (ICF) is a World Health Organization (WHO) framework for measuring health and disability at both individual and population levels that is designed to supplement diagnostic labels to redress the shortcomings of diagnostic labels (WHO, 2001). The ICF is founded on the biopsychosocial model of functioning, which suggests that we must understand how a given condition affects a person's life by considering the intermeshed influences of the biological, psychological and societal consequences of having that condition (Rauch, Lückenkemper, & Cieza, 2012). Accordingly, the ICF aims to conceptualise a person's level of functioning as the outcome of a complex interaction between health, body functions and structures, activities and participation, environmental and personal factors. The interaction among these components, within this account, are dynamic and bidirectional.

A working group employed a rigorous, iterative process to link the ICF concepts to the functional outcomes associated with autism. To do this, a systematic review of the literature (de Schipper et al., 2015), a qualitative exploratory study (Mahdi et al., 2018), an online survey of 225 autism experts (de Schipper et al., 2016) and an international meeting of 20 autism experts (Bölte et al., 2018) were conducted. These studies collectively selected 164 candidate categories from the wider ICF framework and identified them as being most relevant to autism. These categories were then separated into three developmentally appropriate core-sets for (i) pre-school children, (ii) school-age children and young people and (iii) older adolescents and adults (Bölte et al., 2014; Gan, Tung, Yeh, Chang, & Wang, 2014). It is noteworthy that the working group here consisted of experts in research methodology and clinical practice rather than those with lived experience of being autistic, which may have resulted in different appraisals of the functional outcomes in autism, for example, including the sensory environment in the adult core set. Despite these developments, no user-friendly questionnaire to assess these autism-specific ICF core-sets is yet available. The World Health Organisation Disability Assessment Schedule, second edition (WHODAS-2; Garin et al., 2010) has been devised as a brief general measure of self- or informant-rated disability to be used across conditions within the ICF framework.

Autism as a deficit, difference or both.

Autism was first described by, and remains under the control of, medical communities and so a medical-model slant that considers autism as a 'deficit', 'disease' or 'disorder' to be cured or prevented has been traditionally endorsed, either implicitly or explicitly. Autism's place in medical diagnostic manuals has however grown increasingly uncomfortable with repeated calls to adopt a social model of disability that recognises that disability is caused, not by individual differences in ability, but rather by societal responses to those who are different (Hacking, 1999; Oliver, 2013). An intermediate position, termed the social relational model of disability (C. Thomas, 2004), is that the extent to which a person is disabled or enabled involves a complex interplay between the biological and cognitive differences they have and the societal response, or lack thereof, to those differences.

The Autistic community have increasingly challenged and shifted the notion of how autism ought to be understood (Bagatell, 2018). One such activist, Judy Singer (1999) coined the term 'neurodiversity' to draw a parallel between the importance of diversity for thriving social environments and the use of the word biodiversity to highlight the importance of biological diversity for thriving physical environments. Importantly, a large US survey showed that, among autistic respondents, ascribing to the neurodiversity paradigm was associated with whether somebody conceived of their own autism as a positive difference, a deficit or both (Kapp, Gillespie-Lynch, Sherman, & Hutman, 2013).

The neural correlates of autism

Autistic people's brains are different from the brains of non-autistic people. We are still learning precisely how they are different, but some evidence has accumulated in recent years about the differences in structure and in function.

Brain structure. Studies employing structural magnetic resonance imaging show, that young children on the autism spectrum have increased brain volume compared with non-autistic 2-4 year olds, mostly in frontal and temporal regions, which disappear by adolescence, followed by atypically rapid decline in adulthood (Carper, Moses, Tigue, & Courchesne, 2002; Courchesne, 2002; Courchesne et al., 2011; Lange et al., 2015). Both excessive (Thomas, Davis, Karmiloff-Smith, Knowland, & Charman, 2016) and reduced (Frith, 2004) synaptic pruning have been proposed to explain these differences.

With respect to grey matter, a large-scale study found increasing left lateralised cortical thickness in autistic people from age 6 years onwards, which diminished with age, and correlated with symptom severity (Khundrakpam, Lewis, Kostopoulos, Carbonell, & Evans, 2017). With respect to white matter, diffusion tensor imaging has shown reduced structural integrity of long range brain connections (Aoki, Abe, Nippashi, & Yamasue, 2013; Shukla, Keehn, Smylie, & Müller, 2011) but increased integrity in some short range connections, for example, within the corpus callosum (Wolff et al., 2015). White matter integrity appears to be less linked with age among autistic adolescents in the right superior longitudinal fasciculus, a region associated with flexible behaviour (Lisiecka et al., 2015). Localised structural differences in autism are wide ranging, including in the dorsolateral prefrontal, medial-frontal, orbitofrontal, inferior-frontal, anterior cingulate, and posterior parietal cortices; the superior temporal sulcus and fusiform gyrus; the amygdala, hippocampus, thalamus, basal ganglia; and in the cerebellum (for reviews see, Amaral, Schumann, & Nordahl, 2008; and Ecker, 2017).

Brain function. Studies in which participants undergo functional magnetic resonance imaging, electroencephalography, magnetoencephalography and functional near-infrared spectroscopy while completing motor, visual processing, EF, auditory and language and social cognitive tasks have revealed patterns of activation in the brains of autistic participants that differ from non-autistic participants (O'Reilly et al., 2017; Philip et al., 2012). Results from functional neuroimaging studies are heterogeneous but, taken together, tend to suggest that subtle modulation of brain function in response to changes in task demands, under-activation in response to social stimuli and reduced recruitment of the prefrontal regions typically employed to complete EF tasks are characteristic of autism. One of the most reliable findings is that intra-participant and trial-by-trial variability is greater in autistic compared with non-autistic participants (Haigh et al., 2016; Milne, 2011).

Cognitive theories of autism

The heterogeneity in findings about the biological underpinnings and characterising behaviour in autism have made linking the two a complex, often unsuccessful task. Motivated to bridge the gap between brain and behaviour, Frith, Morton and Leslie (1991) suggested that cognition may be the elusive node needed to tie the assorted findings from the disparate research disciplines together. A huge amount of research emerged aiming to uncover the singular cognitive atypicality that could explain the manifold behaviour seen in autism and act as a halfway point or endophenotype between autistic behaviour and its supposed biological cause. An in-depth review of all cognitive accounts of autism is beyond the scope of this thesis (for a review of the classic theories see, Happé, 1994; and for a review of the empirical literature see, Pellicano, 2012; Rajendran & Mitchell, 2007) but the most prominent accounts are introduced briefly below.

Theory of Mind. In brief, theory of mind (ToM) refers to the ability to impute a mental state to oneself and to other people (Premack & Woodruff, 1978). Cognitively able autistic adults struggle, relative to non-autistic adults, to attribute emotional states to others based on information from their eyes alone (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997). While autistic adults pass explicit false-belief tasks such as the Sally-Anne task, they show differences in anticipatory looking patterns on single trial (Senju, Southgate, White, & Frith, 2009) and multiple trial (Schneider, Slaughter, Bayliss, & Dux, 2013) analogues of these classic tasks, suggestive that an implicit ToM difficulty persists into adulthood. Although, this difference attenuates when participants have more experience (Schuwerk, Vuori, & Sodian, 2015) and recently failed to replicate (Schuwerk, Priewasser, Sodian, & Perner, 2018), possibly undermining the legitimacy of implicit ToM assessments.

Weak central coherence. The weak central coherence account purports that autism is the result of a different cognitive style that focuses on local details at the expense of processing the gestalt (Happé & Frith, 2006; Happé, 1996). A recent review suggests that there is neither enhanced local visual processing nor a deficit in global visual processing but instead autistic people appear to be slower to engage in global processing when compared with non-autistic people (Van der Hallen, Evers, Brewaeys, Van den Noortgate, & Wagemans, 2015). Such a finding is consistent with the theoretical position that espouses that global local processing is not mandatory in autism, but rather, "the default setting of autistic perception is more locally oriented than is that of non-autistics" (Mottron, Dawson, Bertone, & Wang, 2007, p. 578).

Other cognitive accounts of autism. Another suggestion has been that autistic people have attenuated Bayesian priors and so rely less on past experience when making sense of current sensory input (Pellicano & Burr, 2012). A biological extension of this account suggests that autism results from a disordered ability to predict what will happen next, which if true, could explain social and non-social difficulties experienced by autistic people as well as explicating why some people become overwhelmed by sensory information (R. P. Lawson, Rees, & Friston, 2014). Alternatively, autism could result from an over-functioning of neural networks typically underlining primary perceptual functions and this over-functioning may also explain social and non-social diagnostic features (Mottron et al., 2006; Mottron & Burack, 2001). Autistic people have been shown to be able to tolerate a higher perceptual load before performance on a secondary task is impacted by increased difficulty in a central task suggesting that autistic people have a higher perceptual capacity (Remington, Swettenham, Campbell, & Coleman, 2009; Remington, Swettenham, & Lavie, 2012) than non-autistic people.

Executive Function

EF difficulties have been proposed as another cognitive explanation for autistic behaviour, it is this account that is the focus of this thesis. I will briefly outline early theories of frontal lobe functions, the subcomponents that are grouped under this umbrella term and the reasons why EF has received so much attention in the literature, in turn below, before reviewing the literature about EF and autism.

Early theory. Interest in the frontal lobe functions that would later become subsumed under the EF umbrella began with the work of Luria (1966). Luria (1966) was the first to describe the disorganisation of actions and strategy application in groups of frontal lobe lesion patients. Norman and Shallice (1986) extended this early view by demarcating the role of two complementary but distinct supervisory systems. On this account, the contention schedule system controls the prioritisation and implementation of routine behaviours and the supervisory attentional system regulates non-routine and novel tasks. The supervisory attentional system, which broadly aligns with modern conceptions of EF, becomes active, according to Norman and Shallice (1986), in situations:

> (a) That involve planning or decision-making; (b) that involve error correction; (c) where responses are not well-learned or contain novel sequences of actions; (d) where danger is anticipated; and (e) which require the overcoming of a strong habitual response or resisting temptation (pp. 21-22).

An alternative account was that optimal human behaviour involves formulating, storing and error checking lists of goals and sub-goals (Duncan & Owen, 2000). On this view, goal-oriented behaviour was thought to be facilitated by a unitary construct, a view borne out of findings that injury to frontal brain regions resulted in relatively global goal-neglect (Duncan, 1986). This goal-neglect construct was believed by its proponents to overlap with, or even be a subsidiary of, general intelligence (Duncan, Johnson, Swales, & Freer, 1997). Stuss and Benson (1986) instead proposed that EF was a fundamental aspect of their tripartite model of frontal lobe function whereby it governed planning, response selection and performance monitoring. Another early theoretical account that laid the groundwork for understanding frontal lobe functions was the somatic marker hypothesis (Damasio, 1995). According to Damasio, people with damage to the ventromedial frontal cortices struggle to regulate their behaviour because they are unable to mark inappropriate or incorrect behaviour with a somatic marker and consequently lack the emotion-related cues needed to organise behaviour.

It was Baddeley and Hitch's (1974) now celebrated multi-component model of working memory that brought the supervisory component of cognition, referred to as the 'central executive', squarely into the remit of cognitive psychology. While undeniably ground-breaking, the central executive was also poorly specified, Baddeley himself referred to it as a "conceptual ragbag" (1996, p. 6). As such, Baddeley and Hitch were unable to answer the question of what was supervising the central executive without drawing on an infinite hierarchy of executives or homunculi, and, consequently, calls have been made to give the term a dignified retirement (Logie, 2016). More recently, the focus has shifted toward understanding the interactive relationship between the cognitive component parts of EF. Accordingly, EFs are a collection of top-down mental processes that develop and are applied in concert and collectively allow behaviour that is in the service of a goal, particularly with respect to novel tasks for which performing the task 'on automatic' would be ill-advised, insufficient, or impossible (Diamond, 2013; Miller & Cohen, 2001). It is important to note that, while a global definition of EF exists and evidence suggests that very early in development a global EF ability exists (Visu-Petra, Cheie, Benga, & Miclea, 2012), this ability becomes increasingly modularised through development (Testa, Bennett, & Ponsford, 2012) and is no longer considered a singular ability, at least later in development. Rather, EFs are a selection of component abilities, and so EF difficulties may exist in a single component, in a constellation of components or globally, across all components.

Subcomponents of EF. The ways in which we ought to carve EF into component parts is the subject of fierce debate (Banich, 2009). One approach to deciphering the componential structure of EF has been to assess participants on a wide-ranging battery of EF tasks, determining which tasks share considerable variance and then creating latent EF variables using statistical techniques such as factor analysis. A widely endorsed model, employing precisely this methodology, found evidence for three overlapping, yet independent constructs: (i) inhibition – the active suppression of some previously activated cognitive representation, (ii) updating and monitoring of working memory – the ability to maintain and manipulate information and (iii) set-shifting – the ability to switch between cognitive sets (Miyake et al., 2000). The core subcomponents of EF together facilitate the completion of complex executive tasks such as problem solving, planning, and navigation (Diamond, 2013), see a schematic of this model in Figure 1.

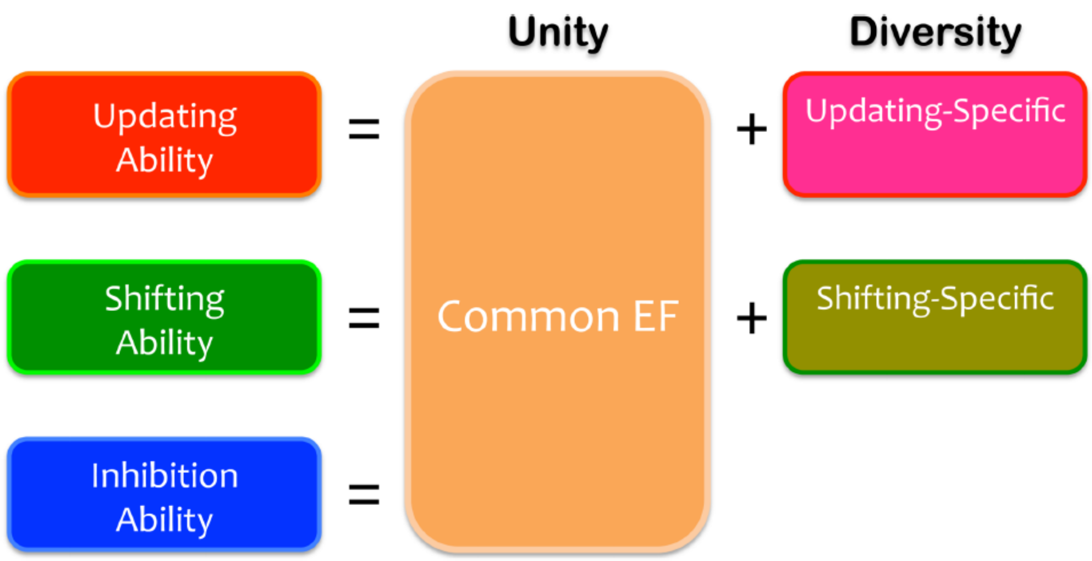


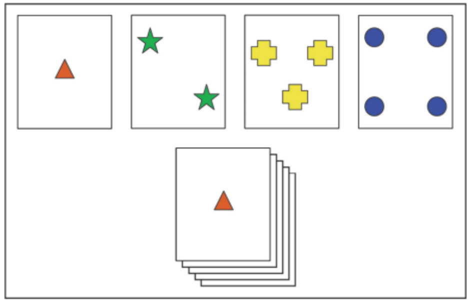
Figure 1. The interrelationship between subcomponents of executive function, according to Miyake and (2012). Figure reproduced from Miyake et al. (2012).

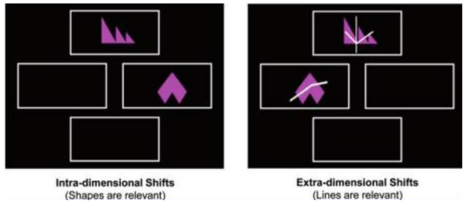
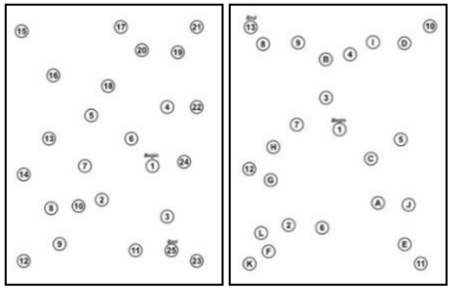
Another distinction made, draws on the somatic marker hypothesis (Damasio, 1995), and refers, on the one hand, to the components of EF that involve affective and motivational drives as "hot" EF. "Cool" EF, on the other hand, refer to cognitive abilities that are more purely cognitive and rational, to be completed without an emotional appraisal (Zelazo & Müller, 2002).

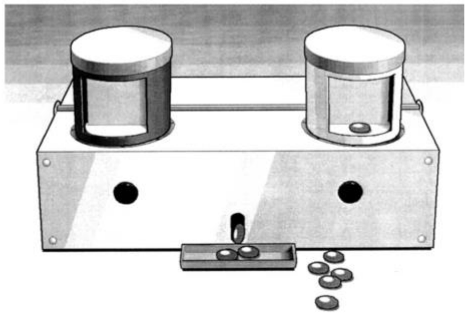
Neuropsychological tasks that measure EF.

The measures employed to assess EF depend on the theoretical paradigms with which they are aligned. For example, assessments based on Luria's (1966) conceptualisation of EF included the following component parts: planning, anticipation, execution and self-monitoring. Assessments based on this conceptualisation involve the assessment of neurological signs of frontal lobe lesions, such as difficulties with motor coordination, sensory integration and disinhibition (Chen et al., 1995). No formal assessments have been based on the goal-neglect theory, largely because global goal neglect is so readily apparent without nuanced assessment (R. C. K. Chan, Shum, Toulopoulou, & Chen, 2008). Traditionally, the most prolific neuropsychological assessments of EF employed by researchers and clinicians fall within the tripartite model of frontal lobe functioning and involve managing conflicting information, flexibly switching between aspects of a task or activities that require sustained control over attentional resources. Examples include the Wisconsin Card Sorting Task (WCST), the Go-NoGo task and the Stroop task; see Table 1 for a description of commonly-used EF assessments.

Table 1.
Commonly used neuropsychological measures of executive function in autism research

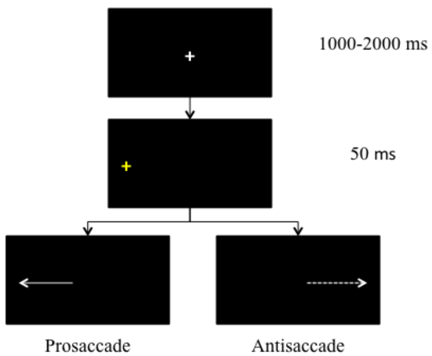
	Sample stimuli	Description	Original reference	Autism studies
Cognitive Flexibility				
Wisconsin Card Sorting Task (WCST)		The participant must sort a deck of 128 cards to four stacks of target decks of cards according to the colour, shape, or quantity of geometric shapes on the card. After the participant correctly matches ten cards, the examiner changes the sorting rule unbeknownst to the participant.	**WCST** (Grant & Berg, 1948)	**WCST** (for a meta-analysis, see Landry & Al-Taie, 2016)
		Variants *The Modified Wisconsin Card Sorting Task (MCST)*: The change in sorting rule is made explicit and only 48 response cards are used, such that cards can only ever be matched to the target deck by one sorting criteria to remove ambiguous feedback.	**MCST** (Nelson, 1976)	**MCST** (Barnard, Muldoon, Hasan, O'Brien, & Stewart, 2008; Hill & Bird, 2006)
		Teddy-bear shifting task: Modified for preschool age children. Participant must identify whether a given card is one of a teddy bear's favourite cards. There are two dimensions to sort by (colour and shape). A new deck of cards with different shapes and colours and a new teddy bear are used after the rule changes.	**Teddy Bear set-shifting task** (Hughes, 1998)	**Teddy Bear set-shifting task** (Pellicano, 2007, 2010)
		Dimensional change card sort task (DCCS): Conceptually similar to the teddy-bear set-shifting task, children are required to sort a series of bivalent test cards, first according to one dimension (e.g., colour), and then according to the other (e.g., shape). This differs from the teddy-bear set-shifting task as there is no teddy-bear used to enhance the salience of the task to young children and because the	**DCCS** (Zelazo, 2006)	**DCCS** (Dichter et al., 2010; Durrleman, Franck, Stephanie, & Julie, 2015; Faja & Dawson, 2014; Pellicano et al., 2017; Zelazo,

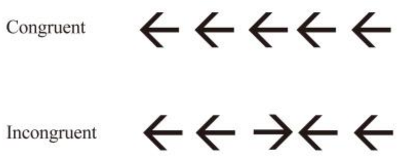
		examiner does tell children that the rule has been changed unlike in the teddy bear task.		Jacques, Burack, & Frye, 2002)
CANTAB Intra-Dimensional/Extra-Dimensional (ID/ED) Shift Task		The ID/ED task begins with simple stimulus discriminations and after six correct responses the rule for determining the correct stimulus changes. The initial intra-dimensional shift involves simple visual discrimination between objects within the same dimension, deciding for example, between two white lines that differ in shape. The extra-dimensional shift then involves shifting attention from one dimension (e.g., lines) to a previously irrelevant dimension (e.g., an overlaid shape). Both the intra- and extra-dimensional shift conditions involve a reversal rule, where participants must maintain the same rule but choose alternate stimuli exemplars. The ID/ED shift task overcomes the shortcomings in the WCST task in that it allows for the separate assessment of cognitive flexibility, maintenance, stabilisation of representations and sensitivity to distraction.	(Cambridge Cognition, 2002; Owen, Downes, Sahakian, Polkey, & Robbins, 1990)	(Corbett, Constantine, Hendren, Rocke, & Ozonoff, 2009; Goldberg et al., 2005; Happé, Booth, Charlton, & Hughes, 2006; Landa & Goldberg, 2005; Ozonoff et al., 2004; Sinzig, Morsch, Bruning, Schmidt, & Lehmkuhl, 2008; Yerys et al., 2009)
The Trail Making Task		In *Part A,* participants must join 25 sequential numbered circles together as quickly and as accurately as they can. In *Part B,* participants must again join 25 circles in sequential order quickly and accurately, but this time must alternately join circles containing letters and numbers (e.g., 1-A-2-B-3-C) **Variants** *The D-KEFS Trail-Making Task:* The administration is the same but normative data are provided. *The Children's Color Trail-Making Task*: A modification for children	**D-KEFS** (Delis, Kaplan, & Kramer, 2001a)	(Bölte, Duketis, Poustka, & Holtmann, 2011; Corbett et al., 2009; Goldstein, Johnson, & Minshew, 2001; Goldstein, Minshew, Allen, & Seaton, 2002; Hill & Bird, 2006; Lopez, Lincoln, Ozonoff, & Lai, 2005; Losh et al., 2009; Minshew,

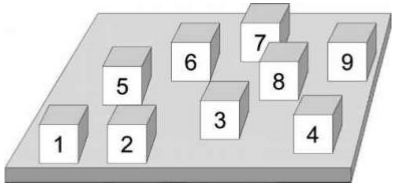
		where even numbers are embedded within yellow circles and odd numbers within pink circles. In *Part A*, participants must join alternate numbers. In *Part B,* a duplicate of each number is included, and each number is embedded both within a pink and yellow circle, participants must join the circles in ascending order but alternate between yellow and pink circles.	**Children's Color Trail Test** (Llorente, Williams, Satz, & D'Elia, 2003; J. Williams et al., 1995)	Meyer, & Goldstein, 2002; Rumsey & Hamburger, 1988; Towgood, Meuwese, Gilbert, Turner, & Burgess, 2009) **Children's Color Trail Test** (Corbett et al., 2009)
Inhibition				
Means-end tasks		Tasks that assess participants' ability to follow arbitrary means to achieve a certain end, typically to win sweets or some rewarding prize. **Variants** *Windows task:* A child is told that there would always be a chocolate in one of two presented boxes, but that neither the child nor a confederate taking part would know which box it was in. The child was asked to point to one box for the confederate, their competitor in this game, to check for a chocolate. If the confederate found the chocolate, they kept it, if not the child won it. Then boxes with one transparent side, facing the child was introduced. The child could then win the chocolate on each trial if they could inhibit the desire to point toward the reward. *Automated windows task:* conceptually the same as the windows task but the procedure is automated, as in the picture. *Less is more task:* A simplified means end-task where two transparent boxes are held in front of a child, one containing a large amount of a given reward, the other containing a smaller	**Windows** (Russell, Mauthner, Sharpe, & Tidswell, 1991) **Less is more** (Carlson, Davis, & Leach,	**Windows task** (Bíró & Russell, 2001; Hughes et al., 1997; Russell et al., 1991; Yerys, Hepburn, Pennington, & Rogers, 2007) **Automated windows task** (Russell, Hala, & Hill, 2003) **Less is more task** (Pellicano et al., 2017) **A not B** (Griffith, Pennington, Wehner, & Rogers, 1999; McEvoy, Rogers, &

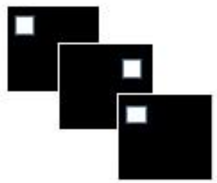
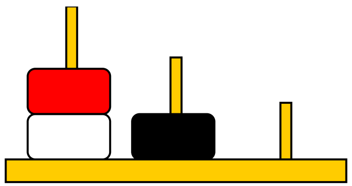
		amount. Children must point at a less desired reward (e.g., 2 stickers) in order to receive a more desired reward (e.g., 5 stickers). *A not B task*: A young child is shown a desired object hidden in one of two locations, after successfully finding it from one location on successive attempts; it is then hidden in an alternate location and they must inhibit looking in the previously correct location.	2005) **A not B** (Piaget, 1954)	Pennington, 1993; Russell, Jarrold, & Hood, 1999)
Stroop Task		Participants are first asked to read aloud and as fast as they can, the names of colours that are printed either in a colour that is congruent or incongruent with the colour named by the word. The difference in time taken to read the list of words is thought to measure inhibition of distracting information.	(Stroop, 1935)	(Adams & Jarrold, 2009; Ames & Jarrold, 2007; Eskes, Bryson, & McCormick, 1990; Goldberg et al., 2005; Jahromi, Bryce, & Swanson, 2013; Johnston, Madden, Bramham, & Russell, 2011; Ozonoff & Jensen, 1999; Ozonoff & Strayer, 1997; S. Robinson, Goddard, Dritschel, Wisley, & Howlin, 2009; Russell et al., 1999; Xiao et al., 2012)
		Variants: *D-KEFS Stroop:* administration is identical to above; normative data are provided.	**D-KEFS** (Delis, Kaplan, & Kramer, 2001b)	**D-KEFS** (Corbett et al., 2009; Semrud-Clikeman, Walkowiak, Wilkinson, & Butcher, 2010)

		Day/night Stroop: adapted to remove the need for participants to be literate. Instead participants must say the word 'day' when presented with a picture of a night sky and vice versa.	**Day/night** (Gerstadt, Hong, & Diamond, 1994)	**Day/night** (Joseph, Mcgrath, & Tager-Flusberg, 2013)
Go/No-Go task		Participants are asked to press a button when they see a target object on screen (go target) and to withhold pressing any button when a non-target object appears (no-go target). Participants are asked to respond as quickly and as accurately as they can. The go target is shown much more frequently than the no-go target and for a period at the beginning of the task, to build up a pre-potent pattern of responding that participants must then inhibit. **Variants** *Stop-signal task (SST):* The go target is present on all trials and participants must wait for a signal before responding. On a minority of trials after the respond signal is presented a different stop signal is presented and participants must withhold their response. *CANTAB stop-signal task (SST):* The participant must press an arrow key to indicate whether a stimulus is an arrow pointing to the left or right. In the second part, the participant is instructed to withhold their response on trials when an audio tone is present. *The continuous performance task (CPT):* Participants are asked to respond (typically via button-press) as quickly as they can when a given stimulus is presented (e.g., a geometric shape on screen) and to withhold a response when any different but similar stimuli appear.	(Donders, 1969; Verbruggen, Logan, & Stevens, 2008) **SST** (Logan, 1994) **CANTAB SST** (Cambridge Cognition, 2002) **CPT** (Rosvold, Mirsky, Sarason, Bransome,	(Adamo et al., 2014; Brandimonte, Filippello, Coluccia, Altgassen, & Kliegel, 2011; Christ, Holt, White, & Green, 2007; Geurts, Begeer, & Stockmann, 2009; Happé, Booth, et al., 2006; Kana, Keller, Minshew, & Just, 2007; Langen et al., 2012; Lee et al., 2009; Ozonoff, Strayer, McMahon, & Filloux, 1994; Raymaekers, Van Der Meere, & Roeyers, 2006; Raymaekers, Antrop, Van Der Meere, Wiersema, & Roeyers, 2007; Sanderson & Allen, 2013; Schmitz et al., 2006; Sinzig et al.,

			& Beck, 1956)	2008; Xiao et al., 2012) **SST** (Adams & Jarrold, 2012; Geurts, Verté, Oosterlaan, Roeyers, & Sergeant, 2004; Lemon, Gargaro, Enticott, & Rinehart, 2011; Ozonoff & Strayer, 1997) **CPT** (A. S. Chan et al., 2009; Corbett & Constantine, 2006; Corbett et al., 2009; Kilinçaslan, Mukaddes, Küçükyazici, & Gürvit, 2010)
Anti-saccade task		Participants are asked to fixate on a cross in the centre of a screen and in the *pro-saccade* condition, they are asked to move their eyes to toward the cross that appears in line with the central fixation point on the left or right periphery of the screen. In the *anti-saccade* condition, participants are asked to move their eyes in the opposite direction from the side of the screen where the cross appears.	(Hallett, 1978)	(Agam, Joseph, Barton, & Manoach, 2010; DeSouza, 2002; Goldberg et al., 2002; Mosconi et al., 2009; Pierrot-Deseilligny et al., 2003)

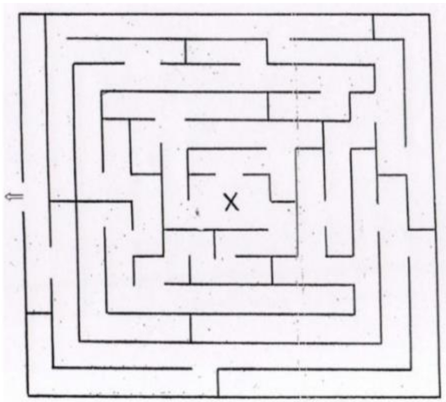
Flanker task		Participants are required to provide a response as quickly as they can either in the presence of information that promotes a correct response (congruent condition) or in the presence of information designed to induce an incorrect response (incongruent condition).	(Eriksen & Eriksen, 1974)	(Adams & Jarrold, 2012; Burack, 1994; Christ et al., 2007; Christ, Kester, Bodner, & Miles, 2011; Dichter & Belger, 2007, 2008; Geurts, Luman, & Van Meel, 2008; Henderson et al., 2006; Iarocci & Burack, 2004; South, Larson, Krauskopf, & Clawson, 2010)
Working Memory				
Verbal span tasks	1. 5 9 0 2. 4 8 6 1 3. 7 3 0 9 4 4. 2 4 9 6 5 8 5. 1 4 6 8 2 4 5 6. 3 9 2 1 5 7 6 0 7. 6 2 5 7 3 9 1 8 4 8. 0 6 3 8 9 4 1 7 2 5	The *forward condition* of the digit span task typically involves an examiner calling out a sequence of digits and asking participants to recall the digits in order. Some variants present the digits on a monitor or by play an audio recording. The number of digits (set size) is incrementally increased until a stop rule is reached. In the *backward condition* of the digit span task, the examiner calls out a sequence of digits as before, but participants are required to recall them in a reverse sequence, starting with the last uttered digit and finishing with the first. **Variants** *Word span task:* Words are used instead of digits *Sentence span task:* Full sentences are read out instead of digits *Non-word span task:* Non-words are read out – this is typically used a control condition in the word span task.	(Wechsler, 1998)	(Andersen et al., 2015; Cui, Gao, Chen, Zou, & Wang, 2010; Gabig, 2008; García-Villamisar & Della Sala, 2002; Gonzalez-Gadea et al., 2013; Ham et al., 2011; Joseph et al., 2013; Joseph, Steele, Meyer, & Tager-Flusberg, 2005; Nakahachi et al., 2006; Russell, Jarrold, & Henry, 1996; D. L. Williams, Goldstein,

				& Minshew, 2006; Zinke et al., 2010)
Spatial span tasks		Participants must remember the spatial location of a given stimuli presented in different locations in a sequence of increasing length and recount them either in the exact same order or in reverse order, as in the verbal span tasks.	**Spatial span** (Wechsler, 1998)	**Spatial span** (Crane, Goddard, & Pring, 2013; Geurts & Vissers, 2012; D. L. Williams, Goldstein, Carpenter, & Minshew, 2005; D. L. Williams, Goldstein, & Minshew, 2005)
		Variants *CANTAB spatial span task:* White squares are shown on a screen, some of which briefly change colour in a sequence of incremental length (ranging from two to nine). The participant must select the boxes which changed colour in the same order that they were displayed by the computer (for the forward variant) or in the reverse order (for backward variant).	**CANTAB spatial span** (Cambridge Cognition, 2002)	**CANTAB Spatial span task** (Goldberg et al., 2005; Happé & Frith, 2006; Landa & Goldberg, 2005; Sachse et al., 2013; Steele, Minshew, Luna, & Sweeney, 2007)
		Corsi blocks: An examiner places an array of blocks on front of the participants and taps the blocks with their finger in a sequence of incremental length and participants must tap the blocks in the same or in the reverse order that the examiner tapped them in.	**Corsi blocks** (Lezak, 1983)	**Corsi blocks** (Englund, Decker, Allen, & Roberts, 2014; Geurts et al., 2004; Pellicano et al., 2017; Verté, Geurts, Roeyers, Oosterlaan, &

				Sergeant, 2005; Zinke et al., 2010)
		Self-order pointing task (SOPT): A number of coloured squares are shown on a screen and participants must find a token hidden behind one box through a process of elimination by tapping on the box to temporarily reveal whether it is occluding a token. The number of boxes is gradually increased until a maximum of 12 boxes are shown.	**SOPT** (Petrides & Milner, 1982)	**SOPT** (Geurts et al., 2004; Joseph et al., 2005; Verté, Geurts, Roeyers, Oosterlaan, & Sergeant, 2006; Verté et al., 2005)
N-back		In a *spatial n-back*, participants are presented with a stimulus (e.g., a square) appearing in one of a fixed number of locations on screen and must decide if the location of the stimuli is the same as the location it was presented *n* steps earlier in the sequence. The task can be made more difficult by increasing the *n* to increase the load on working memory. In a *verbal n-back*, participants are shown a sequence of words, letters or numbers and asked to indicate if a target stimulus is the same or different to the stimuli presented *n* trials ago.	(Kirchner, 1958)	(Barendse et al., 2017; Cui et al., 2010; De Vries & Geurts, 2014; Lever, Werkle-Bergner, Brandmaier, Ridderinkhof, & Geurts, 2015; Rahko et al., 2016; Urbain, Pang, & Taylor, 2015; D. L. Williams, Goldstein, Carpenter, et al., 2005)
Planning				
Tower tasks		The participant is presented with a board containing three pegs of incremental height and three beads that differ in colour that are placed in a particular configuration (the start state) on the pegboard. The examiner then shows the participant a picture of the beads in a different configuration (the end condition) and asks the participants to move the beads to the end condition in as few	**ToL** (Shallice, 1982)	**Tower of London** (Geurts et al., 2004; Kimhi, Shoam-Kugelmas, Agam Ben-Artzi, Ben-Moshe, &

moves as possible and by moving only one bead at a time. Between two and five moves are required to create the end state.		Bauminger-Zviely, 2014; Limoges, Bolduc, Berthiaume, Mottron, & Godbout, 2013; Panerai, Tasca, Ferri, Genitori D'Arrigo, & Elia, 2014; Pellicano, 2007, 2010; Pellicano, Maybery, Durkin, & Maley, 2006; S. Robinson et al., 2009; Schurink, Hartman, Scherder, Houwen, & Visscher, 2012; Unterrainer et al., 2016; Verté et al., 2005; D. Williams & Jarrold, 2013; Zinke et al., 2010)
Variants		
Tower of London –Drexel edition (ToL^{DX}): Six and seven move problems are introduced to increase the sensitivity to finer grained difficulties. The examiner creates the end state with identical pegboard and beads.	**ToL^{DX}** (Culbertson & Zillmer, 1998)	**ToL^{DX}** (Geurts & Vissers, 2012; Wallace et al., 2016)
Tower of Hanoi: Participants are presented with a board containing three pegs of equal length and a number of disks or ascending size are placed on one peg. Participants are required to move the discs, one a time, without placing it off the pegboard and by only placing smaller disks on top of larger disks. The number of disks can be increased to increase the task difficulty.	**ToH** (Borys, Spitz, & Dorans, 1982; Piaget,	**ToH** (Bölte et al., 2011; Griebling et al., 2010; Hanson & Atance, 2014; Keary et al., 2009; Losh et al., 2009; Medeiros

	1978)	& Winsler, 2014; Ozonoff & Jensen, 1999; D. L. Williams, Mazefsky, Walker, Minshew, & Goldstein, 2014)
The D-KEFS Tower of California: Similar to above except each trial increases in difficulty up to a possible 26 moves and the examiner records time taken to complete first move, time to complete the tower, whether the tower was completed correctly or not, number of rule violations. A total achievement score is calculated from there and normative data are available.	**Tower of California** (Delis et al., 2001b)	**Tower of California** (Lopez et al., 2005; McCrimmon, Schwean, Saklofske, Montgomery, & Brady, 2012; Semrud-Clikeman et al., 2010; Van Eylen, Boets, Steyaert, Wagemans, & Noens, 2015)
The CANTAB Stockings of Cambridge (SoC): A computerised version of a tower task in which a participant is shown two displays where three stockings, one containing three coloured balls, is suspended from a beam in each display. Participants must move the balls in one display, one a time, in order to recreate the pattern in the second display.	**SoC** (Cambridge Cognition, 2002)	**Stockings of Cambridge** (Corbett et al., 2009; Goldberg et al., 2005; Happé & Frith, 2006; Hughes, Russell, & Robbins, 1994; Kaufmann et al., 2013; Landa & Goldberg, 2005; Ozonoff et al., 2004; Sachse et al., 2013; Sinzig et al., 2008)
The NEPSY Tower: Twenty trials of increasing difficulty modelled	**NEPSY**	

		off the original Tower of London task with standardised administration instructions and normative data are available.	**Tower** (Korkman, Kirk, & Kemp, 2007)	**NEPSY Tower** (Joseph et al., 2013; Planche & Lemonnier, 2012)
Maze tasks		Participants were asked to complete a series of progressively more complex mazes. To complete a maze, participants had to plan their route ahead to allow them to arrive at the opening of the maze without making any errors. A participant is marked as making an error each time they deviated from the correct path.	(Wechsler, Golombok, & Rust, 1992)	(Barnard et al., 2008; Kimhi et al., 2014; Liss et al., 2001; Low, Goddard, & Melser, 2009; Pellicano, 2007, 2010, 2013; Pellicano et al., 2006; Prior & Hoffmann, 1990)

Assessing discrete subcomponents of EF with tasks such as those described in Table 1 has become widespread among neuropsychological researchers because these assessments can be administered with relative ease and because they are informative in uncovering which neural structures and associated skills have been spared and which have not, post head injury. Similarly, tasks that maximally distinguish between subcomponents of EF have allowed us insight into functional independence of some abilities, or the fractionation of EF. While informative in advancing the basic science of EF, the fractionation of EF is problematic with respect to clinical assessments. That is, a person's performance on a task measuring one subcomponent of EF may not be very informative about how they will perform on a task measuring a different subcomponent of EF and may be less informative still about how they would perform in a complex real world environment (Burgess, 1997; Burgess, Alderman, Evans, Emslie, & Wilson, 1998). For this reason, there has been increasing attention given to measuring EF within the framework of the supervisory attentional system by incorporating more complex, multifaceted and real-life challenges that tap a number of executive domains concurrently (Burgess, 2000b, 2000a). I will review tasks that have applied this approach to the assessment of autistic people in more detail in Chapter 3.

EF has powerful explanatory power. One reason the neuropsychological assessments of EF have received a great deal of attention in the research literature is because they can be powerful predictors of important life outcomes, often over prolonged developmental periods. For instance, robust correlations between EF and ToM have been shown in typical development (Devine & Hughes, 2014). Tests of the well-known 'marshmallow test' demonstrated that the cognitive control needed to delay gratification in preschool children was correlated with academic achievement measured in adolescence (Mischel, Shoda, & Peake, 1988; Shoda & Mischel, 1990; but, see Watts, Duncan, & Quan, 2018). Broader measures of EF have been shown to be related to the extent to which a child is ready to begin school (Blair & Raver, 2015), the emergence of mathematic and literacy abilities in the first year of schooling (Blair & Razza, 2007) and academic achievement over a longer time span (Best, Miller, & Naglieri, 2011). In fact, parent-ratings of their children's self-control ability predicted physical health, criminal offending rates, addiction and personal finances, independent of the role of IQ or social class over a 32-year period, underscoring the predictive power of EF (Moffitt et al., 2011).

EF interventions. Another reason for the attention garnered by EF is that it has a protracted period of postnatal development, making it a prime target for intervention. Widespread development of 'brain-training' games were developed on this premise: games that target the improvement of EF of, for the most part, school-aged children or older adults experiencing the deleterious effects of cognitive aging, through deliberately practising executively demanding games. The initial excitement about the utility of brain games appear to have been unfounded, with little convincing evidence that improvements in the trained task transfer to minimally or especially to distantly related tasks (for a review, see Simons et al., 2016). For instance, one recent study showed that a commercially available brain-training app had no effect on brain activity, cognitive ability or decision making (Kable et al., 2017). There is, however, mounting evidence that diverse activities such as school curricula, in early childhood and martial arts and mindfulness, at any age, have positive impacts on the development of EF (Diamond, 2014; Diamond & Lee, 2011; Diamond & Ling, 2016), meaning that identifying and intervening with EF difficulties is a worthwhile endeavour. One such successful example, is a school curriculum built on the work of Lev Vygotsky called Tools of the Mind, which has been shown to improve the development of cognitive control in children (Blair & Raver, 2014; Diamond, Barnett, Thomas, & Munro, 2007). Interestingly, the largest effect sizes were found for those children from an economically disadvantaged area, and who had weaker EF at the outset (Blair & Raver, 2014; Raver, Blair, & Willoughby, 2012), suggesting those at an early executive disadvantage may stand to gain the most from EF intervention.

Autism as a 'Dysexecutive syndrome'

The EF of autistic people came into focus when an autistic man completed a neuropsychological assessment battery and revealed marked difficulties with problem-solving and showed rigid, perseverative behaviour (Steel, Gorman, & Flexman, 1984) similar to that seen in frontal lesion patients, described as having a dysexecutive syndrome (DES; Damasio & Maurer, 1978). Patients with DES, following lesions to the mesolimbic cortex, neostriatum, and thalamus, have diminished abilities to plan, sequence and execute simple day-to-day activities, engage in perseveration and show a lack of impulse control despite a relative sparing of their general intellectual abilities (Baddeley & Wilson, 1988). The initial account put forward was that autism might be a developmental analogue to the acquired difficulties seen in patients with DES, that is, the cognitive and behavioural atypicalities seen in autism, such as a diminished propensity for joint attention, imaginative play, RRBs, ToM and social communication and interaction, were all secondary consequences of a primary EF deficit (Russell, 1997).

Russell's proposition was that the ability to monitor your own intentions and act with volition were necessary precursors to understanding the intention of any action, be that an action you or another person make (Russell, 2013). According to Russell (2013), it was the executive component of this ability to understand the intentions behind one's own actions, and by association, the ability to understand others, that was at the core of the difficulties observed in those on the autism spectrum. The theory of autism as a primarily DES quickly grew to prominence because, unlike the ToM hypothesis (Baron-Cohen, Leslie, & Frith, 1985), it could account for non-social, as well as, social features of autism. I provide a brief overview of the literature of EF in autistic people below.

The neuropsychological EF assessment of autistic people

The possibility that a single cognitive atypicality could explain the development of autistic behavioural traits led to a proliferation of studies comparing the EF of autistic and non-autistic people. Initial investigations employed assessments that were commonly used in the neuropsychological assessment of lesion patients to assess, for example, cognitive flexibility, inhibition, working memory or tasks assessing higher-order executive abilities such as planning. Investigations of EF in autism range with respect to their theoretical approach and the paradigms of assessment employed. Some investigations assess multiple subdomains to determine whether there is a specific pattern of capability and difficulty for autistic people, other investigations involve the assessment of a single domain as a proxy for overall EF ability or to test a specific hypothesis. Below, I review, in turn, the literature on the cognitive flexibility, inhibition, working memory and complex EF tasks before covering findings pertaining to overall considerations of EF.

Cognitive flexibility. Cognitive flexibility is a person's facility to switch selectively between mental processes in order to produce appropriate behaviour in response to changes in the environment (Dajani & Uddin, 2015). Cognitive flexibility is measured in a variety of ways and is typically operationalised by the tasks used to measure it, meaning that cognitive flexibility can refer to attentional shifting, cognitive set shifting or behavioural task switching ability. A quantitative review of 72 studies published between 1980 and 2012, assessing cognitive flexibility with these types of measures, demonstrated that autistic people have, on average, more cognitive flexibility difficulties than non-autistic people (Leung & Zakzanis, 2014). The authors found the largest effect sizes on the Wisconsin Card Sort Task (WCST) and that overall, autistic participants had greater difficulties on computer-administered versions than they did in traditional, experimenter-administered versions, despite earlier counter evidence (e.g., Ozonoff, 1995). While some metrics yielded large effect sizes, no performance-based measures of cognitive flexibility reliably distinguished between autistic and non-autistic participants and the magnitude of difficulties varied considerably across tests and samples involved. As such, cognitive inflexibility is not universally found in autistic people, although additional possible moderators such as co-occurring conditions, medication use and different matching techniques were not accounted for by Leung and Zakzanis (2014).

The WCST has been so widely used in autism research that a meta-analysis could be conducted on 31 studies comparing autistic and non-autistic children and adults on the task over a 30-year period (Landry & Al-Taie, 2016). The authors found medium-to-large effect sizes (Cohen's *d* = 0.31-0.82) for the four measures of performance typically reported: sets completed, perseveration, failure-to-maintain-set, and non-perseverative errors. For autistic participants, Performance IQ and age predicted perseverative error rates whereas Verbal IQ predicted non-perseverative error rates. Nearly one third of studies reporting on number of sets achieved and number of perseverative errors generated null findings and more than half of the studies reporting failure-to-maintain-set and non-perseverative errors generated null findings (Landry & Al-Taie, 2016).

Dependent variables have been inconsistently reported in WCST studies and experimental measures are often insufficiently standardised, making it difficult to interpret the extent to which researcher degrees of freedom influence the pattern of results and whether a multiple comparisons problem has been invoked via a "garden of forking paths" (Gelman & Loken, 2014). Furthermore, the average sample size in WCST studies was 25 participants per group, allowing only for the reliable detection of effects of large magnitudes and leaving uncertainty about whether difficulties are absent in some samples or, instead, if the studies seeking to answer this question are underpowered (Landry & Al-Taie, 2016). Effect sizes appear to diminish with increasing age, but samples were inconsistently or poorly matched, particularly in earlier studies.

Robust group differences have also been found on measures of cognitive flexibility adapted to be developmentally appropriate such as set-shifting tasks (Dichter et al., 2010; Pellicano, 2007, 2010; Pellicano et al., 2017; Zelazo et al., 2002), a trail-making test (Corbett et al., 2009) and the inhibition switching subtest of the Developmental Neuropsychological Assessment (Narzisi, Muratori, Calderoni, Fabbro, & Urgesi, 2013). The intra-dimensional/extra-dimension (ID/ED) shift task, which allows researchers to control for differences in participants' ability to learn from feedback, to retain a task goal and to sustain a consistent response over time, is less likely to show group differences than classic set-shifting paradigms (Landa & Goldberg, 2005; Ozonoff et al., 2004; Yerys et al., 2009).

The conclusion often drawn within the literature is that limitations in cognitive flexibility cause the behavioural inflexibility at the core of autism, such as difficulties shifting flexibly between tasks or adopting the perspectives of other people (D'Cruz et al., 2013). This conclusion has strong face-validity but it has been called into question because the evidence for cognitive flexibility difficulties from robust experimental measures such as the ID/ED is inconsistent and because there is a large gap between everyday difficulties in flexible behaviour and the type of set-shifting measured by neuropsychological tasks (Geurts, Corbett, & Solomon, 2009). One recurrent suggestion for the inconsistent pattern of results is that neuropsychological assessments lack ecological validity, and that performance on these assessments may not map onto everyday behavioural inflexibility (Geurts, Corbett, et al., 2009; Hill, 2004a; Landry & Al-Taie, 2016).

It is difficult to interpret performance on a single task such as the WCST or the ID/ED shift task as necessarily representing difficulties with cognitive flexibility, given the multifactorial nature of these measures, which have been criticised for involving many more abilities than just cognitive flexibility (Geurts, Corbett, et al., 2009; Landry & Al-Taie, 2016). In order to discern reliably whether cognitive flexibility difficulties putatively cause behavioural inflexibility, Geurts and colleagues (2009) argue, we must expand the 'toolbox' of measures used to measure flexibility, especially to include tasks with ecological validity and those that account for the role of non-executive factors. I return to this issue in Chapter 3.

Inhibition. Inhibition involves the following subcomponents, exerting control over impulses and desires to engage in goal-directed behaviour (self-control), suppressing pre-potent mental representations (cognitive inhibition), the ability to supress a previously activated motor response (pre-potent response inhibition) and the ability to attend to one aspect of a task stimulus while ignoring another (resistance to perceptual distractors).

With respect to self-control, the 'windows task' showed that autistic children struggled to engage in strategic deception to win a motivating reward visible through a window, such as, for example, chocolate, by failing to misdirect the examiner toward a box they know does not contain the reward (Russell, Mauthner, Sharpe, & Tidswell, 1991). To ensure that performance on this task was likely the result of failing to disengage from a salient object rather than a weakness in ToM ability, a follow-up study removed the interpersonal competition within the task and reproduced the findings (Hughes & Russell, 1993). In a similar way, the Less-is-More paradigm (Carlson et al., 2005), has shown autistic pre-schoolers struggle with conflict inhibition (Pellicano et al., 2017). Similarly, cognitively able autistic and non-autistic children were compared on the classic marshmallow task (Mischel et al., 1988) and autistic children were less likely to delay gratification in order to receive the larger reward (Faja & Dawson, 2013).

On the Stroop task, autistic participants tend to perform as well as non-autistic participants, unlike those with ADHD who consistently struggle (Eskes et al., 1990; Ozonoff & Jensen, 1999; Ozonoff & Strayer, 1997; Russell et al., 1999). There are however two issues with using the Stroop task in autism. First, it measures both resistance to distractors and pre-potent response inhibition within the same task and so it is unclear if those who struggle on the task have difficulties with one or both cognitive skills, or if the simultaneous application of two skills is what they find most difficult. Second, for autistic participants, the size of the interference effect has been shown to be related to the extent to which an individual has reading comprehension difficulties (Adams & Jarrold, 2009). Notwithstanding, autistic children have been shown to have difficulties on a day-night Stroop task, which does not require reading comprehension (Russell et al., 1999).

There have been many examples in the literature demonstrating that autistic people struggle to inhibit attention to distractors (Ames & Jarrold, 2007; Christ et al., 2007, 2011; Corbett & Constantine, 2006; Corbett et al., 2009). There are nevertheless exceptions to these studies, where autistic people perform comparably to non-autistic people (Goldberg et al., 2005; Schmitz et al., 2006; Solomon, Ozonoff, Cummings, & Carter, 2008; Xiao et al., 2012).

With respect to the inhibition of pre-potent responses, autistic participants have been shown to under-perform relative to non-autistic people (Bishop & Norbury, 2005; Christ et al., 2007; Corbett et al., 2009; Geurts et al., 2004; Langen et al., 2012; Xiao et al., 2012). Although almost as many studies have also demonstrated no significant group differences (A. S. Chan et al., 2009; Happé & Frith, 2006; Ozonoff et al., 1994; Schmitz et al., 2006; Sinzig et al., 2008). Indeed, when individual differences in processing speed, general cognitive ability and age were accounted for, no group differences were found on measures of pre-potent response inhibition or proactive interference but differences remained on the flanker task (Christ et al., 2007, 2011). Similarly, autistic participants have shown intact ability to inhibit a pre-potent response on a stop-signal task but they did perform more poorly relative to non-autistic participants on a modified Flanker task (Adams & Jarrold, 2012).

Findings such as these highlighted the possibility that inhibition in autism might be best characterised as a specific difficulty with inhibiting attention to interference effects. Two separate meta-analyses were conducted, which together included 41 studies, to assess precisely this possibility and found that both pre-potent response inhibition ($d = 0.55$) and interference control ($d = 0.31$) difficulties were observed in autistic people (Geurts, van den Bergh, & Ruzzano, 2014). The slightly larger effect sizes for pre-potent response inhibition tasks runs contrary to previous evidence (Christ et al., 2007, 2011). Age moderated performance on pre-potent response inhibition tasks such that increasing age was related to a decrease in the observed effect size across studies but did not moderate effects in interference control tasks, contradicting any developmental effects in meta-analyses that have considered all inhibitory control tasks together (Demetriou et al., 2018; E. C. L. Lai et al., 2017). IQ moderated effect sizes in interference control tasks such that higher IQ was related to lower observed effect sizes across studies but did not moderate effects for pre-potent response inhibition tasks. Consequently, results may differ across tasks depending as a function of sample matching criteria. Variability between study effects remained considerable, after controlling for inhibition type, age and IQ, suggesting additional factors were moderating the effects.

Bíró and Russell (2001) suggested another explanation for the heterogeneity in findings: that autistic people struggle specifically with following arbitrary procedures rather than necessarily with inhibition. To test this possibility, they conducted a study containing a task with arbitrariness but without an element of pre-potency, a task with both attributes and a task with neither attribute and the results supported their hypothesis that arbitrary procedures were the source of group differences. Moreover, the computerised assessment of response inhibition achieved reduced effect sizes compared with in-person assessments (Demetriou et al., 2018). Together, these results suggest that autistic people as a group tend to underperform on measures of inhibition but that their performance is impacted by many factors that are not specific to inhibition or necessarily to EF.

Working memory. Baddeley defined working memory as the "brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning" (1992, p. 556). The workings of this brain system are generally studied with respect to three component parts of Baddeley's model of working memory: (i) the phonological loop - the ability to rehearse and store language information; (ii) the visual sketchpad - the ability to hold and manipulate visual information in mind; and (iii) the central executive - the supervisory attentional aspect.

Bennetto et al. (1996) conducted a study to determine if autistic people have a specific difficulty in the processes involved in storing or manipulating linguistic information (the phonological loop), often referred to as verbal working memory. The authors compared the performance of cognitively able autistic adolescents with a clinical comparison group matched on sex, age and VIQ on the counting span and the sentence span tasks. The autistic participants performed more poorly than the non-autistic participants, leading to further studies in which autistic participants were shown to have difficulties in similar verbal working memory span tasks (Gabig, 2008; Ham et al., 2011; D. L. Williams et al., 2006). Despite these investigations, other investigations have found intact working memory ability in pre-schoolers (Griffith et al., 1999), children and adolescents (Ozonoff & Strayer, 2001) and adults (Geurts & Vissers, 2012) on the autism spectrum.

To test the integrity of the processes required to maintain and to manipulate non-linguistic, spatially represented information (i.e., the visuospatial sketchpad), often referred to as spatial working memory, one task often used is the Spatial Working Memory task in the CANTAB battery (Robbins et al., 2010). A study comparing autistic and non-autistic children and young adults on this task found autistic people struggled relative to their non-autistic peers, indicating a difficulty with spatial working memory (Steele et al., 2007). Similar spatial span tasks have found group differences between autistic and non-autistic groups (Geurts et al., 2004; Geurts & Vissers, 2012; Joseph et al., 2013; Sinzig et al., 2008; D. L. Williams, Goldstein, Carpenter, et al., 2005; D. L. Williams et al., 2006). Although, like with verbal working memory tasks, there are exceptions where spatial working memory in autism appears to be intact (Dawson et al., 2016; Edgin & Pennington, 2005; Lopez et al., 2005; Ozonoff & Strayer, 2001).

When verbal and non-verbal working memory are assessed within the same autism sample, difficulties with spatial working memory but not verbal working memory are found (Ambery, Russell, Perry, Morris, & Murphy, 2006; D. L. Williams, Goldstein, Carpenter, et al., 2005). Another suggested explanation for discrepant findings is that autistic people struggle with complex but not simple working memory tasks (Nakahachi et al., 2006). Of note here is that there was a correlation between language ability and working memory for non-autistic children but no such relationship for autistic children, suggesting poor performance was caused, not by language difficulties, but by a failure to use language in the service of task performance (Joseph et al., 2013). Notwithstanding, autistic children appear to be as likely as non-autistic children to engage in articulatory rehearsal, despite having a reduced central executive capacity (Russell, Jarrold, & Henry, 1996).

A meta-analysis of 28 studies comparing the working memory of autistic and non-autistic participants showed that, in fact, autistic participants struggled with both spatial and verbal working memory, although larger effect sizes were found for spatial (d = 0.72) compared to verbal (d = .44) tasks (Wang et al., 2017). Moreover, this meta-analysis showed autistic participants had a comparable amount of difficulty on maintenance trials (where they were required to hold blocks of information in mind and to recount them immediately) and on manipulation trials (where they were required to hold blocks of information in mind and apply a manipulation to the information, such as recounting it in reverse order). Age and IQ moderated the observed effect sizes, such that increasing age and IQ were associated with a reduction in effect sizes across studies but, after controlling for these, there was significant heterogeneity in results suggesting further moderators were implicated, but not assessed (Wang et al., 2017). A more nuanced picture emerged from another meta-analysis that considered childhood, adolescence and adulthood separately and found group differences in childhood and adulthood but not in adolescences, owing, according to the authors, to reduced working memory ability in typical adolescents possibly as a consequence of neural reorganisation that is caused by the onset of puberty (Demetriou et al., 2018).

Planning, multitasking, abstract problem solving. In addition to tasks that aimed to measure single subcomponents of EF in as pure a form as could be achieved, there have been investigations that sought to measure activities that require the application of multiple component executive skills in pursuit of a goal, so-called higher order EF. One such example that has received much attention in the autism literature is planning. Early investigations of the planning ability of autistic people tended to employ the Tower of London task and consistently found autistic participants struggled to complete the task (Hughes et al., 1994). Subsequently, several variants of this task, including the Drexel edition of the Tower of London, the Tower of Hanoi and the Stockings of Cambridge (from the CANTAB battery) have been employed with autistic people.

Some studies do not find differences in performance on measures of planning between autistic and non-autistic people (e.g., Bölte et al., 2011) while others find poorer planning performance among autistic participants (e.g., Brunsdon et al., 2015). A number of suggestions have been made to explain the discrepant findings in planning ability in autism, such as the heterogeneity in symptom severity, task type, or non-executive factors such as age or IQ (Hill, 2004b; Kenworthy, Yerys, Anthony, & Wallace, 2008). Relatedly, the gap between autistic and non-autistic participants in everyday functional behavioural skills, commonly referred to as adaptive skills, such as self-care and independent goal setting has been found to increase from childhood to adolescence and adulthood (Pellicano, Cribb, & Kenny, 2018; Pugliese et al., 2015, 2016). Interestingly, childhood performance on a tower task predicted significant variance in a group of cognitively able young autistic adults' adaptive behaviour 12-years later (Kenny, Cribb, & Pellicano, 2018), suggesting that planning difficulties become increasingly salient as autistic adolescents have increasing demands placed on them. Unlike cognitive flexibility, but similar to inhibition tasks, autistic people perform worse on the standard human-administered neuropsychological tasks such as the Tower of London task (Lopez et al., 2005) than on computer-administered variants such as the CANTAB Stockings of Cambridge (for a review see, Kenworthy et al., 2008). Also, IQ has been seen to be more strongly related to performance on planning measures in autistic than in non-autistic people, supporting the suggestion that IQ may be playing a different role in planning tasks for autistic people (Brunsdon et al., 2015).

A meta-analysis of 50 planning studies in autism showed a significant medium positive effect (Hedges' $g = 0.52$), indicating that autistic people perform worse on planning tasks compared to non-autistic samples (Olde Dubbelink & Geurts, 2017). As expected, there was significant heterogeneity in effect sizes across planning studies (p <

.0001). The meta-analysis did not support the suggestion that planning difficulties increase in adolescence but found that difficulties persist at a consistent rate right across development. This is consistent with another meta-analysis that found evidence for the stability of planning difficulties among autistic people across development (Demetriou et al., 2018). The widening gap between autistic and non-autistic participants in adaptive behaviour as they transition from childhood through adolescence and into adulthood is likely not a result of growing planning difficulties but rather a consequence of an increasing mismatch between their planning ability and the extent to which independent planning is required. According to Dubbelink and Geurts (2017), age, IQ or task type did not significantly moderate the effect sizes across the included studies but the authors suggest that the choice of comparison group, co-occurring conditions and the use of psychotropic medication among participants may impact the meta-analytic findings presented, but that these were inconsistently reported in individual studies. There was also evidence of a publication bias, whereby null findings have likely gone unpublished.

Overall EF ability. A meta-analysis of 235 studies conducted between 1980 and 2016 comparing the EF ability of autistic and non-autistic people (aged 6 years and above) found that autistic people had moderate EF difficulties and had small-to-moderate effect sizes in each of the EF subdomains included in their analysis (Demetriou et al., 2018). Interestingly, there was no significant difference in the magnitude of the effect sizes between different domains suggesting that autistic people, on average, struggled relative to non-autistic people on every EF subdomain assessed. Demetriou and colleagues suggest, on balance, that EF difficulties in autism are global and are not fractionated by subdomain, which together with the largely linear trajectory of EF development, are broadly explained by under, and/or over-connectivity between brain networks. With respect to developmental effects, there were no group differences in working memory, fluency and cognitive flexibility during adolescence despite apparent difficulties in each of these domains in childhood and adulthood. Overall, only working memory effect sizes were significantly moderated by age. Year of publication had a small but significant moderating effect whereby more recently published studies typically had smaller effect sizes, possibly reflecting the increasing variability in cognitive ability of participants.

Another meta-analysis, this time including ninety-eight studies, by Eric C. L. Lai and colleagues (2017) focused specifically on the potential role of individual differences in intellectual abilities and co-occurring ADHD in studies focusing on children and adolescents. When participants who were closely matched on intellectual ability and who did not have an ADHD diagnosis were considered, Lai et al. (2017) found that autistic participants had moderate difficulties in flexibility, generativity, and both verbal and spatial working memory. Although contrary to Demetriou et al. (2018), autistic and non-autistic participants did not differ on inhibition ability and the group differences on planning tasks only remained when samples differed with respect to ADHD symptomatology and intellectual ability. Autistic participants had greater difficulties in generativity than did those with a dual diagnosis. Like Demetriou et al. (2018), a meta-regression showed that older autistic participants had fewer difficulties on measures of cognitive flexibility than younger participants did, suggesting that autistic adolescents show a developmental spurt in their cognitive flexibility skills as they learn and increasingly demonstrate flexibility skills, as they get older. It might also be that non-autistic adolescent's cognitive flexibility skills decrease in adolescence because of the neural reorganisation that occurs during puberty or it might instead be that tasks become increasingly insensitive to difficulties with increasing age.

The overarching view of the literature pertaining to the EF of autistic people is marked with inconsistency and rife with contradictory evidence. Synthesising the literature is challenging because of the diverging or underspecified theoretical accounts of EF being assessed, the variety of assessments used and the lack of standardised assessments to allow for cross-study comparisons. These issues are often compounded by sample sizes that are insufficient to detect anything but the largest possible effect sizes and by investigations that conflate performance on a single measure of a given executive construct with that construct, rather than creating latent variables of executive subcomponents.

Evidence against EF as causal in the development of autism

Despite initial interest and a proliferation of studies designed to assess whether aberrant EF development was a contributory, causal factor in the development of autism, the evidence has not supported this position. Specifically, the five criteria for determining the causality of EF difficulties in autism, namely that they are (i) universal, (ii) specific, (iii) related to diagnostic behaviour (iv) have causal precedence and (v) can subsume other cognitive accounts, have not been met (Ozonoff, Rogers, & Pennington, 1991; Pennington et al., 1997). I briefly outline the reasons why below.

Universality. One method that could demonstrate whether EF difficulties are a central and primary part of the cognitive profile in autism would be to demonstrate that such difficulties are universal among (all or almost all) autistic people. In support of this account, one early assessment found that 96% (n = 23) of the cognitively able autistic children and adolescents assessed had difficulties, demarcated by achieving an executive composite score below the mean of the non-autistic group (Ozonoff, Rogers, & Pennington, 1991). Similarly, Ozonoff and Jensen (1999) showed that 39 out of 40 autistic participants had difficulties on at least one measure of EF, that is, they scored below the mean score of the non-autistic sample for each respective measure. Hill and Bird (2006) showed that 21 out of 22 autistic participants were identified as having difficulties on EF, but this time the criteria used to determine a difficulty was that their score fell below the 5th percentile of the control sample on at least one executive measure. Pellicano (2007), however, failed to find evidence of universality and showed that 33%, 43% and 50% of autistic children (aged 4 – 7 years) fell one standard deviation or more below the mean of the non-autistic sample on experimental measures of inhibition, planning and cognitive flexibility, respectively.

Categorising 'impairment' is somewhat arbitrary as is evident in the variety of criteria used in each of the above-mentioned studies. Nevertheless, it is clear from Pellicano (2007) and the variety of studies that did not find group differences on a range of classic EF measures, when variance in non-executive factors such as IQ and psychomotor processing are controlled for (e.g., Hill & Bird, 2006; Liss et al., 2001), that not all autistic people have difficulties with every EF task. These findings rule out a universal global EF difficulty as an explanation for autistic behaviour. It remains possible, however, that autistic people universally have difficulties in a given subdomain or pattern of subdomains of EF.

Specificity. If a global EF impairment had a specific one-to-one mapping with autistic behaviour, then it should not be possible for EF difficulties to exist in the absence of characteristically autistic behaviour. Yet EF difficulties without autistic-like behaviour are present in other developmental populations including, for example, Tourette's syndrome (Channon, Pratt, & Robertson, 2003), schizophrenia (Hutton et al., 2017), ADHD (Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005) and cognitive aging (Robbins et al., 1998). The fact EF difficulties do not automatically result in a syndrome-specific pattern of behaviour has been termed the 'discriminant validity problem' (Burgess et al., 1998). While the discriminant validity problem might be damaging for the possibility that an EF difficulty could explain the development of autistic behaviour, it is nevertheless possible that an autism-specific difficulty in one EF subcomponent, or a particular constellation of EFs, could explain some autistic features. For example, Ozonoff and Jensen (1999) suggested performance on measures of cognitive flexibility and planning could discriminate between autistic and non-autistic participants and those with other conditions.

Sinzig and colleagues (2008) found, on the one hand, that children with ADHD had difficulties with inhibition and working memory tasks and, on the other hand, autistic children had difficulties with planning and flexibility tasks. Whereas those with an autism and a with co-occurring ADHD diagnosis were similar to the autistic participants who did not have ADHD with respect to inhibition but not working memory ability. A comparison between adults with ADHD and autism found the ADHD group had difficulty withholding a response, showed preservation in initiation and planning abilities, the autism group, by contrast, exhibited significant difficulties in initiation, planning and strategy formation (Bramham et al., 2009). A review of 28 studies, similarly found that those with ADHD and autism were similar to those with only autism on both flexibility and planning and similar to those with only ADHD on response inhibition (Craig et al., 2016). Children on the autism spectrum had difficulties in all EF domains except interference control and working memory but had comparatively more difficulties than those with ADHD on measures of planning and flexibility (Geurts et al., 2004). Autistic children, when compared with children with non-autistic communication disorders, were found to struggle more on measures of inhibition (Bishop & Norbury, 2005).

Overall, it appears that studies are more likely to yield group differences between autistic and non-autistic samples on measures of flexibility and planning, but there is not sufficient evidence for an autism-specific profile of EF difficulties. Rather, on balance, autistic samples seem to have difficulties on all measures of EF (Demetriou et al., 2018). The discriminant validity problem does therefore undermine the possibility that EF difficulties are causal to the development of autistic behaviour.

Relationship with diagnostic behaviour. Another central tenet of the theory of autism as a dysexecutive syndrome is that EF difficulties ought to be related, statistically, to the autistic behaviours they purport to explain. Set-shifting abilities have been shown to be correlated with repetitive behaviours in preschool children (Yerys et al., 2009) and in adolescents (South, Ozonoff, & Mcmahon, 2007) on the autism spectrum. Interestingly, in another investigation, performance on EF measures were not correlated with autism symptom severity, but was correlated with a measure of adaptive behaviour (Ozonoff et al., 2004). Low levels of effortful control, measured by the marshmallow task, are associated with greater levels of social communication and interaction difficulties in 6-7 year-olds on the autism spectrum (Faja & Dawson, 2013). Response inhibition and task switching were related to motor and sensory stereotyped behaviour (Mostert-Kerckhoffs, Staal, Houben, & de Jonge, 2015). Conversely, however, a delay non-matching to sample task, which taps executive control, was associated with neither autism symptomatology (Dawson, Meltzoff, Osterling, & Rinaldi, 1998) nor joint attention (Dawson et al., 2016) in young autistic children.

While some studies demonstrated significant relationships between working memory and core autistic traits (Landa & Goldberg, 2005; Lopez et al., 2005; Sachse et al., 2013), other studies have not (e.g., Steele et al., 2007). For example, working performance was negatively correlated with problems in communication and social ability (Gilotty et al., 2002). Individual differences in cognitive flexibility, working memory, and response inhibition, as measured by the D-KEFs, were highly related to the level of RRBs engaged in by autistic participants but planning and fluency were unrelated (Lopez et al., 2005). Bölte et al. (2011), by contrast, found that higher order planning difficulties were associated with higher scores for RRBs on the ADI-R and ADOS.

Causal precedence. If EF difficulties lead to the development of autistic behaviour then, according to the causal precedence criterion, EF difficulties should be readily apparent in very young autistic children. One investigation failed to show this however, and instead, demonstrated that the EF of autistic pre-school children (with a mean age of 2.9 years) was no different to chronologically and mentally age matched peers, and so suggested that EF difficulties might be better explained as a secondary consequence of being autistic (Yerys et al., 2007). Another study found that a different cohort of pre-school autistic children did not underperform relative to their non-autistic peers on any of eight EF tasks performed (Griffith et al., 1999). Similarly, 3- and 4-year-old children on the autism spectrum performed comparably to non-autistic children on measures of EF suggesting that, at least at this early age, there is no autism-specific pattern of EF difficulties (Dawson et al., 2016). Similar subsequent investigations reported a lack of group differences (Rutherford & Rogers, 2003; Stahl & Pry, 2002).

Some studies, however, have provided evidence for EF difficulties in very young autistic children. For instance, preschool autistic children have more difficulties on reversal learning activities where they need to reverse a learned rule, such as the location of a hidden sweet, after a number of trials (Dawson et al., 1998; McEvoy et al., 1993). Furthermore, a sample of autistic children aged between three and six years of age had, on average, more difficulties completing the DCCS, the Less is More and the Corsi blocks (backward condition) than non-autistic children matched for age and ability (Pellicano et al., 2017). It remains possible that the tasks employed in the study of EF early in development are not specific to the types of difficulties autistic people have (Bíró & Russell, 2001) or were insufficiently wide ranging to detect difficulties (Pellicano, 2011). Alternatively, these results may suggest EF difficulties emerge later in development than autistic features (cf. Yerys et al., 2007).

Ability to subsume other cognitive accounts. The fifth and final criterion by which cognitive accounts of autism are assessed for developmental primacy is whether it can explain and thus subsume other cognitive accounts (Ozonoff, Rogers, & Pennington, 1991). Longitudinal studies of the explanatory power of EF in autism are best placed to assess this criterion. It has been shown that autistic children between four and seven years of age, can have ToM difficulties while having intact EF, supporting the idea that EF is a forerunner to ToM development (Pellicano, 2007). When some children from the same cohort were assessed three years later individual differences in EF and Weak Central Coherence predicted later ToM ability, but the relationship between EF and ToM was unidirectional, again adding weight to the idea that EF shapes the developmental course of emerging ToM (Pellicano, 2010). In addition, individual differences in EF *but not* ToM predicted later social communication ability and repetitive and restricted behaviours underscoring the developmental primacy of EF not only for cognitive development but also for the expression of diagnostic traits (Pellicano, 2013). Finally, in a recently conducted long-term follow-up of this same cohort, early EF – but not ToM – predicted significant variance in young people's autistic features *and* adaptive functioning 12 years later (Kenny et al., 2018).

Another longitudinal study found both ToM and EF difficulties were present in late childhood (M_{age} = 12 years), persisted over a three year period and followed comparable developmental trajectories (Ozonoff & McEvoy, 1994). The authors concluded that ToM and EF were related but interdependent on each other for development. EF significantly predicted play ability at age 6 years over and above intelligence, but early play did not predict later EF in autistic preschool children with higher language ability (Faja et al., 2016). These results suggest that, at least for verbally able autistic children, early EF abilities may be critical precursors for the development of play. Together, these results demonstrate that rather than there being a distinct pattern of EF atypicalities that relates to specific behavioural features of autism it appears as though early EF ability moulds many aspects of emerging behaviour, including, but not limited to, social behaviour. These atypicalities are likely best understood and studied alongside other cognitive capabilities and atypicalities that have explanatory power over the life outcomes of interest (C. R. G. Jones et al., 2018; Pellicano et al., 2006).

Moving past autism as a 'dysexecutive syndrome'

The evidence from studies of EF in autism presented here demonstrate that individual differences in EF have explanatory power over development for autistic people, as they do in non-autistic people. Nevertheless, EF difficulties in autism are not universal, there are no straightforward ways to discriminate between autistic and non-autistic patterns of EF difficulty, and evidence for EF in very early childhood is, at best, mixed. What is more, EF difficulties are related to autistic features and other cognitive atypicalities in both cross-sectional and longitudinal investigations but measuring multiple cognitive atypicalities and capabilities of autistic people in concert can improve predictive power and, as such, EF does not wholly subsume other cognitive accounts of autism.

In sum, there is no longer sufficient evidence to support the developmental primacy of EF difficulties to the emergence of autistic features, at least, when the strongest version of this theoretical account is pursued. Indeed, the fractionation of autism at the genetic, cognitive and behavioural level has resulted in a shift away from pursuing any single cognitive explanation for the development of autism (Happé, Ronald, & Plomin, 2006). Going forward, research about the EF of autistic people should shift its focus from establishing developmental primacy to developing approaches that can successfully identify individuals whose executive difficulties have real world impacts. Consequently, case-control studies that only have sufficient statistical power to determine whether samples of autistic participants perform substantially more poorly, on average, than their non-autistic counterparts on neuropsychological measures of EF might be of limited value, especially if the relationship between the task and real-world executively-demanding tasks is poorly specified.

Criticism of neuropsychological EF assessments

Cognitive constructs, like EF, are not directly observable and, in lieu of direct measurement, inferences are made about EF ability by recording specific aspects of behaviour produced (e.g., accuracy, reaction time, failure to respond, errors committed) in response to tasks that were designed to test theoretical predictions. Making inferences about EF from behaviour becomes problematic when theories overlap and conflict but are insufficiently specified to allow directly contradictory predictions. Psychological science does not have effective procedures for resolving conflicted theories (Poldrack & Yarkoni, 2016), which is evidenced by a lack of instances when a theory has been completely abandoned in light of a critical experiment (Greenwald, 2012). The result, with respect to EF measurement, is an ever-expanding 'toolbox' of theory-led EF tasks. Synthesising results from tasks derived from different theoretical models, or even from different tasks derived from the same model, is challenging because we lack a formal EF ontology to operationalise when performance on different tasks reflects the same or different theorised constructs or to understand the componential nature of diverse constructs (Amabile, 1983; Poldrack, 2010). Furthermore, referring to tasks as measuring EF has been criticised for being premature, in light of our relative ignorance over what causes poor performance on these tasks and their lack of sensitivity to frontal lesions that should logically be implicated (Rabbitt, 2004). A related issue is that in populations who are 'hard-to-reach' or have limited attentional resources, such as those with developmental disabilities, performance on a single task is conflated with latent constructs, despite cautions against this practice (Cronbach & Meehl, 1955).

The traditional neuropsychological EF tasks borrowed from research with lesion patients have helped identify the presence of EF difficulties in autism, but they are not without their criticisms. Historically, the remit of clinical neuropsychology was limited to identifying the location and extent of brain injuries and abnormalities (Long & Kibby, 1995). To do so, neuropsychologists aimed to evoke the best performance a participant can demonstrate on reliable and valid instruments designed to distinguish maximally between the states of specific brain processes. The role of clinical neuropsychology has, however, evolved toward evaluating the functional consequences of brain-based differences by, for example, informing decisions about a person's fitness to work or their ability to live independently (Heinrichs, 1990).

Initially, these evolving demands resulted in neuropsychologists administering experimental neuropsychological assessments, such as those described in Table 1, which were assumed to be related to the state of specific brain process and *critically* that performance on these measures was informative about how these processes might function beyond the confines of the laboratory (Burgess, Alderman, Evans, Emslie, & Wilson, 1998). Unfortunately, this assumption barely received empirical testing before the practice became relatively mainstream.

To illustrate with respect to EF, Shallice and Burgess (1991) presented case studies of three patients with acquired frontal lobe lesions who performed relatively well on a battery of neuropsychological assessments and consequently were believed to have relatively spared cognitive abilities. Strikingly, each of these people's personal lives had fallen, post head-injury, into disarray. Of particular interest to Shallice and Burgess (1991) was whether there were important differences between the cognitive skills being measured by neuropsychological assessments and the way in which people apply those skills in everyday life. To test this, they developed two novel assessments, the Six Elements Test and the Multiple Errands Test. These assessments, unlike traditional tasks, were open-ended, required participants to apportion their time independently and required sustained attention toward both goal attainment and self-monitoring. All three participants showed significant difficulties with these tasks. This finding, among others, has led to widespread claims that traditional neuropsychological tests, most especially those measuring EF, are insufficiently 'ecologically valid' (Chaytor & Schmitter-Edgecombe, 2003; Manchester, Priestley, & Jackson, 2004).

Definition of ecological validity

Ecological validity was coined by Egon Brunswik (1948) who required a new term when proposing *representative design* to oppose *systematic design* which, at that time, was the overwhelming convention employed within experimental psychological research. Representative design, unlike systematic design, involves designing assessments that represent some attributes of life outside the laboratory. According to Brunswik (1956), just as participants in an experiment must represent those people not in the experiment over whom we wish to generalise the findings, the conditions in an experiment must also represent the conditions outside of the experiment over which we aim to generalise the findings (Hammond, 1998). Both the *substantive material* (for example, recognition of photographs of real faces rather than computer generated ones) presented and the *formal conditions* (e.g., recognition of people rather than static photographs of faces) need to represent those outside the experiment. The systematic approach, by contrast, involves the systematic arrangement of conditions in an experiment to answer a specific question. Systematic design therefore has no concern for ecological validity, which is problematic only if the specific laboratory conditions do not represent the non-systematic arrangement of conditions as they occur outside of the experiment and if somebody wishes to generalise the findings beyond the experiment.

In strict Brunswikian terms, ecological validity refers to how informative a sensory cue is in a study using representative design. For example, by measuring the relationship between a distal cue (e.g., a banana) and the related proximal cue (e.g., the perception of the banana being yellow), a person can make a probabilistic estimate as to the ripeness of that banana. In this example, if the participant is deciding on the ripeness of the banana the ecological validity of the relationship between the distal and proximal cues is close to one, because the correlation between colour and ripeness of a banana is high. If the participant is deciding, instead, how far away the banana is, then the ecological validity of the relationship between cues is approximately zero, because there is no relationship between perceived yellowness and distance from a banana. In an unpublished online essay, Hammond (1998) emphasised the often misguided notion that ecological validity referred to the task being realistic and highlighted that, according to Brunswik, ecological validity can be demonstrated in arbitrary and artificial laboratory tasks:

> The concept of representative design does *not* call for generalization to "real world" conditions (a meaningless demand). Rather, representative design calls for, first, a *specification of the conditions toward which the generalization is intended*, and second, a *specification of how those conditions are represented in the experimental conditions* (para. 7, emphasis in the original).

More recently, however, the concept of ecological validity has been used in a different sense, to refer to the degree to which results obtained under controlled experimental conditions can be inferred to relate to results obtained in uncontrolled, naturalistic environments (Franzen & Arnett, 1997; Spooner & Pachana, 2006; Tupper & Cicerone, 1990), which is how it will be used for this thesis. Ecological validity, in this sense, is comprised of two different, but closely related aspects, representativeness and generalisability. The first, *representativeness*, describes Brunswik's representative design, which is the extent to which something can be investigated in a form and context that corresponds to its occurrence in everyday life. The second, *generalisability*, as the term suggests, refers to the degree to which the results of a particular study are capable of explaining other similar processes or tasks in everyday life (Kvavilashvili & Ellis, 2004). If a study is both generalisable and representative, then it can be said to be ecologically valid. Conversely, if it lacks both of these characteristics, the conclusions made by the study are not ecologically valid.

There is debate about the relative importance of representativeness and generalisability for studies that possess only one such characteristic. A study conducted in an environment outside of the laboratory, for example, can be said to be representatively-designed but it can suffer from weak experimental control, such as, whether a participant uses an aid to achieve the task or how a person not involved in the study behaves toward a participant. Such a lack of control over conditions represents a threat to the internal validity of a study, as performance may be unduly influenced by these factors and so fail to generalise to other related processes in everyday life. This example serves to demonstrate that generalisability is relatively more important than representativeness in determining the ecological validity of a study.

There are two approaches to assess whether a measure has ecological validity; tests of *verisimilitude* – the extent to which generalizable design was used – and tests of *veridicality* – the extent to which performance predicts other, related real-world measures (Burgess et al., 2006, 1998; Chaytor & Schmitter-Edgecombe, 2003; Kenworthy et al., 2008). As a result of the often underwhelming ability to predict functional behaviour from traditional cognitive measures, researchers have called for applying assessments of the veridicality and verisimilitude of EF in typical as well as clinical settings (Spooner & Pachana, 2006).

Ecological validity of the EF assessment of autistic people

The criticisms levelled at traditional neuropsychological measures of EF, such as their lacking ecological validity, have sobering implications for our ability to understand what these measures truly tell us about the abilities and difficulties autistic people have outside of the laboratory. Autistic people often experience profound adaptive difficulties; they are less likely to attain full-time employment, many have fewer friends than they would like, they are more likely to struggle with day-to-day tasks needed to live independently and to participate actively in the community (Howlin, Goode, Hutton, & Rutter, 2004; Howlin, Savage, Moss, Tempier, & Rutter, 2014; Shattuck, Orsmond, Wagner, & Cooper, 2011). These difficulties do appear to abate somewhat through development but, regrettably, most autistic adults do not attain normative outcomes on measures of adaptive skills (Seltzer, Shattuck, Abbeduto, & Greenberg, 2004; Shattuck et al., 2007). The possibility that difficulties with EF could forecast many of these maladaptive outcomes has good face validity but investigations of the EF of autistic people that have empirically tested the ecological validity of the conclusions they espouse are scarce, making such a conclusion premature.

In fact, one particularly illustrative study, with only autistic children who did not have an ID (i.e., an IQ > 70), found that IQ was weakly related to parent-reported functional behaviour (Klin et al., 2007), measured by the Vineland Adaptive Behaviour Scale, second edition (VABS-II; S. S. Sparrow, Cicchetti, & Balla, 2005). Klin et al. (2007) showed adaptive behaviour profiles that were substantially below participants' IQ scores, indicating that cognitively able autistic people often have profound difficulties within adaptive skills that are incommensurate with their cognitive potential. One candidate explanation for the mismatch between performance on standardised cognitive assessments and on assessments of everyday functioning is that some individuals might *test well* in the laboratory or clinic, but nevertheless not necessarily *function well* in everyday life (for a review, see Kenworthy et al., 2008). To test this possibility, a limited number of studies have (i) assessed the EF of autistic people on representatively-designed assessments and (ii) assessed the generalisability of conclusions from EF assessments to other real-world measures of adjustment; these studies will be reviewed in Chapter 3 and Chapter 4, respectively.

Conclusions

The theory that autism is a developmental analogue of the DES, whereby autistic behavioural features emerge as a secondary consequence of divergent executive abilities through development, has not borne out. Irrespective of a failure to support the strongest version of this theory, studies investigating the EF of autistic people have nevertheless remained fervent, but the field of research has arrived at something of an impasse; the ever-expanding literature appears to become ever-more contradictory and difficult to interpret. As we move beyond autism as a dysexecutive syndrome and toward an understanding that individual differences in EF ability are a potentially powerful predictor of the life chances of autistic people, we must be committed to better understanding the relationship between the tasks we employ in the psychology laboratory and the behaviour we claim to explain beyond the laboratory. Understanding how we can take learning from the laboratory to the everyday will be the focus on the remainder of this thesis. To address this issue, I will assess autistic and non-autistic young people on both systematically- and representatively-designed EF tasks to test their relative utility in identifying group differences and in explaining individual differences in EF skills that occur outside of the laboratory.

In Chapter 2, I report on a study assessing autistic and non-autistic adolescents on a wide-ranging battery of EF tasks that are typical of systematically-designed laboratory investigations of EF in autism. Crucially, this battery involves assessing participants on multiple tasks from each executive subdomain so that latent variables can be generated to represent EF ability, reducing the measurement error from non-executive factors so often implicated in the literature. Furthermore, by employing a battery that is open-source, standardised and fully manualised, this study should facilitate future investigations to gain a more nuanced understanding of the interrelationship between EF subcomponents with larger, more diverse samples.

In Chapter 3, I turn to representatively-designed EF tasks, reporting on a study assessing autistic and non-autistic adolescents on such a task that more closely resembles the situations in which EF abilities are deployed in daily life. I further contextualised participants' performance by also accounting for their disposition toward completing the task and in so doing sought to create a better understanding of contextual factors in EF task performance.

In Chapter 4, I test the generalisability of the systematically-designed measure of EF described in Chapter 2 and of the representatively-designed measure of EF described in Chapter 3 to other metrics of other functional abilities, namely participants' self-reported quality of life, everyday EF ability, perceived disability and anxiety.

In Chapter 5, I detail autistic adolescents' and their parents' own perspectives about the development of EF for themselves or their children. In this chapter, I sought to add the voice of young autistic people to the literature about EF and autism and to redress the almost complete absence of the first-person description of executive abilities in autism.

Finally, in the General Discussion (Chapter 6), I make five recommendations on the basis of the research presented within this thesis. That is, by (i) validating systematically-designed tasks against real-world measures, (ii) employing representatively-designed tasks to capture functional ability, (iii) triangulating objective assessment with subjective reports, (iv) conducting carefully controlled experiments to elucidate the mechanisms underlying everyday executive difficulties and (v) conceptualising of EF in autism within a dimensional framework, we can move the discipline of EF research of autistic people forward.

Chapter 2

Comparing the executive function ability of autistic and non-autistic adolescents with a manualised battery of neuropsychological tasks

Introduction

As discussed in Chapter 1, the literature relating to executive function (EF) in autism is extensive, markedly inconsistent and difficult to interpret. In this chapter, I outline some of the difficulties in the measurement of EF that make interpreting the literature challenging and one approach to address these difficulties – the use of an open-source, manualised, EF battery (Kramer et al., 2014). I then describe the first empirical study of my PhD using this battery with cognitively able autistic and non-autistic adolescents. This thesis will focus on cognitively able autistic young people for two reasons. First, the research question being addressed is whether autistic and non-autistic young people who do not have an ID differ with respect to specific cognitive difficulties, namely in EF, which purportedly impact upon their functional skills. Second, we lack valid and reliable measures of EF that are appropriate for individuals with and without an additional ID. Specifically, when assessing EF in those with an ID it is difficult to determine whether performance relates to EF difficulties or to language abilities given that all EF assessments rely on language ability to some extent.

The task impurity problem

EF is an overarching term for a set of theoretical cognitive constructs that can be notoriously difficult to measure. One methodological reason why our ability to make inferences from performance on EF tasks difficult is what has been termed the task impurity problem (Snyder, Miyake, & Hankin, 2015). That is, performance on a single EF task is composed of the systematic variance related to the EF construct of interest and the systematic variance related to EF ability in general, as well as the systematic variance related to non-executive factors (e.g., language ability, motor coordination, reaction time) and non-systematic variance (i.e., error). The predominant approach in autism research has been to use a single task to represent a discrete subdomain of EF ability. On the one hand, this approach could lead one falsely to believe that an EF construct holds strong explanatory power when, in fact, the variance in non-EF or general EF abilities might be driving the effects in which one is interested. On the other hand, it could lead to the false belief that an EF construct holds no explanatory power when, in fact, it does, but we lack tasks to measure the construct with sufficient purity. The impurity problem can be reduced by assessing participants on multiple tasks tapping a given EF construct and creating a latent variable from the shared variance that reflects the EF construct one aims to measure. This issue might be especially pronounced in autism research because few autism studies report on latent measures of EF constructs. Additionally, reviews of standard EF assessments of autistic people show that they are more likely to struggle on multifactorial tasks (e.g. tower tasks and card-sort tasks) that simultaneously probe many EF constructs or measure EF constructs alongside non-executive abilities than they are to struggle on tasks with greater measurement purity (Hill, 2004a; Kenworthy et al., 2008).

Diversity of EF tasks

Another impediment to progress in the EF literature is the diversity of EF tasks employed. The tasks employed are often developed in-house by a single research group, are not formally manualised and are reported upon with insufficient detail to be fully reproducible. Consequently, iterative studies apparently using the same paradigms are tweaked, variables are calculated in new ways, different dependent variables are reported upon and the results cannot be meaningfully synthesised. Potential antidotes to this are (i) using EF measures that are manualised so that administration, scoring and dependent measure selection are standardised, (ii) not using proprietary measures so that identical tasks can be used across research groups and (iii) depositing the resultant data in open repositories to maximise the usefulness of the study to those synthesising evidence across studies.

Small samples

The impact of having a diverse range of EF measures is compounded in autism research by underpowered studies. Original theoretical accounts hypothesised that EF difficulties in autism were substantial and universal (Hughes et al., 1994; Russell, 1997) but, by contrast, findings reported in the literature have been mixed, for instance, autistic participants do not struggle with every measure of EF (e.g., Russell et al., 1999). The expectation that difficulties were substantial resulted in a proliferation of studies that were sensitive only to very large effect sizes. For instance, a recent meta-analysis of the 235 studies comparing EF in autistic and non-autistic samples, aged at least 6 years, conducted between 1980 and 2016 had a median sample size of 23 (Demetriou et al., 2018). Similarly, a meta-analysis of 31 studies comparing autistic and non-autistic samples on the Wisconsin Card Sorting Test between 1980 and 2013 found the median sample size was 25 (Landry & Al-Taie, 2016). Although there is variability in sample sizes in the literature, the median sample sizes described above reflect studies that are powered to detect an effect size of at least 0.8 with power of 0.80 at an alpha of 0.05 (Cohen, 1988; Landry & Al-Taie, 2016). In a review outlining the literature about autistic people's cognitive flexibility, Geurts and colleagues (2009) argued that one important remedy to publishing paradoxical findings is to conduct studies with adequate sample sizes.

As well as facilitating direct replication, open source tasks and data make it possible for many labs to combine EF measurement data, making testing EF scalable, as can be seen from recent collaborative worldwide projects that allowed the assessment of large, diverse samples (Open Science Collaboration, 2015). In future, EF assessment in autism should capitalise on collaborative cross-laboratory studies, to test samples sufficiently large to detect small and moderate effect sizes and sufficiently diverse to test the generalisability of these effects.

The NIH-EXAMINER battery

An open source battery of tasks called Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research (EXAMINER; Kramer et al., 2014) was developed for a National Institute of Health (NIH) funded project. The NIH-EXAMINER battery was designed to integrate the cognitive, experimental and clinical EF literatures to produce a battery of EF tasks with acceptable psychometric properties as well as tested and reported measures of test-retest reliability (ranging from .78 to .93) and validity (Kramer et al., 2014). The NIH-EXAMINER is based on the Miyake et al. (Miyake et al., 2000) model of EF, which found support for three related, but separable, core EF constructs that contribute to complex executive tasks, such as planning. The battery therefore includes task tapping mental set shifting, information updating and monitoring, inhibition of pre-potent responses and planning. Tasks tapping fluency ability have also been included. A confirmatory factor analysis demonstrated that both a one factor (global executive composite) and a three-factor model (fluency, cognitive control and working memory) characterise the data well. Importantly, task administration, scoring and dependent variable selection are completely manualised, and the battery is open source.

The studies that have employed the NIH-EXAMINER battery have predominantly focused on patients with frontal lobe pathophysiology. One such study, showed that patients with behavioural variant Frontotemporal Dementia had comparable difficulties to those with Alzheimer's Disease on the NIH-EXAMINER measures of working memory, semantic fluency and sustained attention but had more difficulties with letter fluency and anti-saccade accuracy, relative to a typical comparison group (Kramer et al., 2014). In this sample, discriminant function analysis could reliably classify participants into diagnostic groups with 73% accuracy (Kramer et al., 2014). The continuous performance task (CPT) from the battery has been used to assess patients with Parkinson's disease pre- and post-treatment in a pharmaceutical intervention study, but did not show change in performance (Possin et al., 2013). Another study of mixed neurological patients found that damage to different neuroanatomical structures associated with EF difficulties was related to performance on different tasks within the battery (H. Robinson, Calamia, Gläscher, Bruss, & Tranel, 2014).

In developmental populations, performance on all individual measures within the NIH-Examiner battery decreased with increasing neurologic morbidity and two factor scores were related to reduced white matter area in a sample of youth with sickle cell disease (Schatz, Stancil, Katz, & Sanchez, 2014). Finally, children and adolescents with attention deficit hyperactivity disorder (ADHD) have been shown to score lower on the working memory but not on the fluency or cognitive control factor scores than those without ADHD (Schreiber, Possin, Girard, & Rey-Casserly, 2014). The battery has not been employed, to my knowledge, to assess EF in autistic participants.

Current study

The aims of the current study were twofold. First, I tested whether autistic adolescents (11 to 19 years) would show, on average, greater difficulties on each task within the NIH-EXAMINER Battery (Kramer et al., 2014) relative to non-autistic participants closely matched on age, gender distribution and intellectual ability. Second, I further assessed whether autistic participants would show, on average, more difficulties on the latent variables generated to represent fluency, cognitive control, working memory and overall EF ability relative to the matched, non-autistic participants.

I predicted on the basis of existing research, that found autistic participants have difficulties on each EF subcomponent (Demetriou et al., 2018; Geurts et al., 2014; E. C. L. Lai et al., 2017; Landry & Al-Taie, 2016; Wang et al., 2017), that, in the current study, the autistic participants would underperform relative to the non-autistic participants on each EF measure. Specifically, I predicted that autistic participants would struggle on each of the individual tasks and on each of the factor scores derived from the NIH-EXAMINER battery. I hypothesised that I would find these group differences because the current study involved samples that were sufficiently large to detect effect sizes of at least a moderate magnitude (in this case, $d > .57$) which is greater than the statistical power in many of the studies detailed in the above-mentioned meta-analyses. To that end, one-tailed hypotheses were used in each comparison.

Finally, I was interested in the extent to which individual differences in non-executive background variables such as chronological age, verbal and non-verbal IQ and symptomatology were related to task performance. I predicted that age and verbal IQ would be significantly positively related to task performance.

Method

Participants

Seventy-nine adolescents aged between 11 and 19 years ($M = 14.76$, $SD = 2.13$) completed every measure in this study, including 37 autistic (12 female, 25 male) and 42 non-autistic (22 female, 30 male) participants. Autistic participants were recruited through an advertised public engagement and research participation event ($n = 25$), community contacts at the Centre for Research in Autism and Education (CRAE; $n = 5$) and a school-research partnership ($n = 7$).

All autistic participants had received an independent clinical diagnosis of autism, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; APA, 2000, or DSM-5; 2013), or The International Classification of Mental and Behavioural Disorders, tenth edition (ICD-10; WHO, 1992)[4] criteria. Participants also scored above the threshold indicative of autism on either the parent-reported Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012) or the Autism

[4] Although The International Classification of Mental and Behavioural Disorders, eleventh edition (ICD-11; WHO, 2018) autism diagnostic criteria are now available, none of the participants in the current study were diagnosed using these criteria.

Diagnostic Observation Scale, second edition (ADOS-2; Lord, Rutter, et al., 2012; data for one participant were available on neither measure). Data for one additional autistic participant were excluded because they obtained an SRS-2 *t*-score of 56, which fell below the threshold strongly associated with having an autism diagnosis (a *t*-score of 60 or above; Constantino & Gruber, 2012; ADOS-2 data were unavailable for this participant).

According to parent report, 20 autistic participants had at least one co-occurring developmental or psychiatric disability. Parents reported that participants had the following conditions; Attention-deficit hyperactivity disorder (ADHD, $n = 9$), dyslexia ($n = 10$), developmental coordination disorder (DCD; $n = 7$), sensory processing disorder ($n = 3$), Irlen Syndrome ($n = 1$), and obsessive-compulsive disorder ($n = 1$). Parents also completed the ADHD symptom checklist (The Centers for Disease Control and Prevention, 2017) and 12 autistic participants were reported to have six or more inattention and hyperactivity symptoms within the last six months, which is suggestive that of a possible ADHD diagnosis, this included five whose parents reported a diagnosis and six whose parents did not. Only one autistic participant did not speak English as their first language.

Non-autistic participants were recruited through an advertised public engagement and research participation event ($n = 40$) or through community contacts at CRAE ($n = 2$). None of the non-autistic participants had any developmental disabilities and one had a diagnosis of depression, as reported by parents. There were three additional non-autistic participants (all female) whose parent-reported SRS-2 *t*-scores were between 60 and 65, above the threshold indicative of an autism spectrum condition and whose data were excluded from all analyses (SRS-2 data were unavailable for ten non-autistic participants). English was not the first language spoken by six non-autistic participants.

There were some noteworthy differences in the distribution of participant ethnicities between groups, reported by parents, including that almost three quarters of autistic participants were from White backgrounds compared with less than half of the non-autistic participants. A Fisher's exact test revealed that these differences were statistically significant, $p = .001$. See Table 2 for more information.

Table 2.
Demographic information about the participants assessed on the NIH-EXAMINER battery

	Non-autistic (n = 42) *n*	**Autistic** (n = 37) *n*
Recruited through		
Public engagement and research participation event	40	25
Community contacts	2	5
School-research partnership	0	7
Parent-reported co-occurring conditions		
Attention deficit hyperactivity disorder	0	9
Dyslexia	0	10
Developmental coordination disorder	0	7
Sensory processing disorder	0	3
Irlen Syndrome	0	1
Obsessive compulsive disorder	0	1
Complex language disorder	0	0
Depression	1	0
Medication		
ADHD	0	5
Antipsychotic	0	1
Sleep inducing	0	4
Antiepileptic	0	1
Antidepressants	1	3
Non-psychoactive	2	2
Ethnicity		
Any White background	13	27
Any Asian background	3	3
Any Black background	9	1
Any mixed background	4	4
Other ethnic group	1	1
Missing or prefer not to say	11	1

Note. Any White background = White British, White Irish or any other White background; Any Black background = Black British, Black African, Black Caribbean or any other Black background; any Asian background = Chinese, Indian, Pakistani, Bangladeshi or any other Asian background; Mixed/multiple ethnic groups = Mixed White and Asian, Mixed White and Black African, Mixed White & Black Caribbean, Any other Mixed background.

All participants were considered cognitively able by virtue of having verbal, nonverbal and full-scale IQ scores of 70 or above (see Table 3). An additional six autistic participants were excluded from the analysis for obtaining IQ scores below this threshold.

Samples were group-matched by age, $t(77) = 0.67$, $p = .502$, $d = 0.15$, verbal IQ , $t(77) = 0.25$, $p = .804$, $d = 0.06$, nonverbal IQ, $t(77) = 0.43$, $p = .671$, $d = 0.10$, and by gender distribution, χ^2 (1, $n = 79$) = 2.43, $p = 0.12$, $\phi = 0.20$.

The socio-economic status (SES) of participants and their parents, indexed by measuring education level, level of neighbourhood deprivation and income and occupation, has been shown to interact with the prevalence of autism diagnoses (Nowell, Brewton, Allain, & Mire, 2015). For that reason, the under-reporting of the SES of participants in autism research has been highlighted as problematic (Delobel-Ayoub et al., 2015).

Number of years of parental/caregiver education is a widely used proxy for SES (Liberatos, Link, & Kelsey, 1987). A parent or caregiver (hereafter, 'parent') reported the age at which they left fulltime education, which was used to derive the number of years of education they received after they turned 16 years of age. There were no significant differences between the autistic and non-autistic participants in level of parental/caregiver education, $t(62) = 0.39$, $p = .699$, $d = 0.10$.

An index of multiple deprivation for each participant was calculated based on the post code reported by parents, using the English Indices of Deprivation 2015 (T. Smith et al., 2015). The English Indices of deprivation provide a numerical value to denote a given neighbourhood's position relative to the national distribution on each of the following criteria; income, employment, education and skills, health and disability, crime, barriers to housing and services, living environment, income deprivation affecting children, and income deprivation affecting older people. An index of multiple deprivation (IMD) combines each of the abovementioned criteria into a single metric. A decile rank IMD was recorded for each participant, accordingly scores ranged from 0 to 10 and lower scores indicated higher levels of deprivation. There was a significant group difference on the decile rank of deprivation, $t(67) = 3.80$, $p < .001$, $d = 0.92$. Years of parental post-16 education and the decile rank of IMD were converted to z-scores and a mean of both was calculated to create a composite SES score. See Table 3 for more. There was a significant group difference on the composite SES score, $t(53) = 0.97$, $p = .339$, $d = 0.26$.

Table 3.
Descriptive statistics for the developmental and background variables of participants assessed on the NIH-EXAMINER battery, by group

	Non-autistic (n = 42)		Autistic (n = 37)		
	M	(SD)	*M*	(SD)	
	Range		Range		*p*-value
Developmental variables					
Gender (M:F)	20:22		25:12		.120
Age (in years)	14.91	(2.25)	14.59	(2.01)	.502
	12 - 18		11 - 19		
Full-scale IQ	103.74	(12.39)	103.78	14.58	.988
	77 - 130		76 - 132		
Verbal IQ	102.98	(13.58)	102.24	(12.42)	.804
	78 - 132		73 - 123		
Nonverbal IQ	103.14	(13.31)	104.73	(19.48)	.671
	71 - 132		75 - 154		
SRS-2	47.96	(5.93)[a]	77.57	(9.32)[b]	**< .001*****
	38 - 59		56 – 90[c]		
ADOS-2 severity score [d]	-	-	5.74	(2.60) [e]	-
	-		2 – 10[f]		
SES of primary caregiver					
Years of post-16 education	5.39	(5.06)[g]	5.00	(2.98)[h]	.699
	-1 – 19[i]		0 - 12		
IMD decile rank	4.89	(3.11)[j]	7.52	(2.45)[k]	**< .001*****
	1 - 10		1 - 10		
Composite SES score	-0.04	(0.78)[l]	0.15	(0.61)[m]	.339
	-1.22 – 1.14		-1.18 – 1.35		

Note. Full-scale IQ = Full-scale IQ, 4 subtest version, derived from the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011), where mean score is 100 and standard deviation is 15; Verbal IQ = Verbal comprehension index derived from the WASI-II (Wechsler, 2011), where mean score is 100 and standard deviation is 15; Nonverbal IQ = Perceptual reasoning index derived from the WASI-II (Wechsler, 2011); SRS-2 = *t*-score, calculated separately, by gender from the Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012); ADOS-2 = the Autism Diagnostic Observation Schedule, second edition (Lord, Rutter, et al., 2012); SES = Socioeconomic status; IMD = Index of Multiple Deprivation (T. Smith et al., 2015); Composite SES score = mean of z-scores derived from the years of post-16 education and the IMD decile rank.
[a] n = 32. [b] n = 35. [c] one participant fell below the threshold indicative of an autism diagnosis (a *t*-score of 60) but were retained in the analysis because they received a score above the threshold indicative of an autism diagnosis on the ADOS-2. [d] The ADOS-2 was only administered with participants already diagnosed with autism. [e] n = 19, The ADOS-2 was not administered to all autistic participants because it was not an included measure in the planned protocol for this study due to time limitations. Data are reported here in cases where ADOS-2 data were available because participants took part in other research studies and consented for their data to be used in this way. [f] Two autistic participants received ADOS-2 severity scores below the threshold indicative of an autism diagnosis (a severity score of 2 or below) but were retained in the analysis because they received a score above the threshold indicative of an autism diagnosis on the SRS-2. [g] n =28. [h] n =36. [i] One parent reported leaving full time education aged 15 years and so received a score of -1 when number of years of post-16 education was calculated. [j] n =36. [k] n =33. [l] n = 23. [m] n = 32.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Measures

The NIH-EXAMINER (memory.ucsf.edu/examiner) tasks are a combination of computer and paper-and-pen measures that tap working memory, inhibition, set-shifting, fluency and planning ability that generate working memory, cognitive control, fluency and global executive composite scores. There are three forms of each task that are of comparable difficulty to allow for repeated testing in the case of either longitudinal or clinical trial research, Form A was used in this study. The computerised tasks within the NIH-EXAMINER were presented with PsychoPy software (Peirce, 2007) and latent factor scores were calculated using item response theory with the *ltm* package (Rizopoulos, 2006) in R (version 3.4.3; R Core Team, 2017) using the script provided by the original authors of the battery (Kramer et al., 2013). The variables that contribute to each factor score and the individual tasks from which these variables are derives are described below.

I calculated a *fluency factor* score, according to manual instructions (Kramer et al., 2013), by combining the total number of correct responses in the semantic and phonemic fluency tasks, which are described below.

Phonemic Fluency. Participants were asked to name as many words as they could that began with a given letter of the alphabet in one minute. Two trials were administered, one with the letter L and another with the letter F. The participant was asked not to give names of places, people or numbers as responses. If participants paused for 15 seconds, they were encouraged to keep going. Correct responses, rule violations and repetitions were scored manually by an examiner, according to the NIH-EXAMINER manual (Kramer et al., 2013). The number of correct responses, rule violations and repetitions were totalled across both trials. A higher number of correct responses was indicative of greater phonemic fluency.

Semantic Fluency. Participants were asked to name as many words as they could within a given category in one minute. Two trials were administered, one for the category 'animals' and another for the category 'vegetables'. Participants were informed that it did not matter which letter the names of vegetables or animals began with. If participants paused for 15 seconds, they were encouraged to keep going. Correct responses, rule violations and repetitions were scored according to the NIH-EXAMINER manual

(Kramer et al., 2013). The number of correct responses, rule violations and repetitions were totalled across both trials. Greater accuracy was indicative of greater semantic fluency.

I calculated a *cognitive control factor score*, according to manual instructions (Kramer et al., 2013), by combining the total shift score from the set-shifting task, the total flanker score from the flanker task, the total score from the anti-saccade condition of the saccade task and the total number of dysexecutive errors recorded by the examiner during administration of the battery.

Set-shifting Task. The set-shifting task was designed to measure cognitive flexibility. On each trial, participants were presented with a target image in the centre of the screen that was either red or blue and either a rectangle or triangle. Participants were also presented with a condition cue (the word shape or the word colour) at the bottom of the screen which indicated which dimension participants should use to match the target to one of the two comparison images presented in the bottom left and right of the screen (see Figure 2). The participant was asked to respond as quickly as possible while avoiding errors. If required to match by colour, the participant was asked to press the left arrow key for red and the right arrow key for blue and, if matching by shape, they were asked to press the left arrow key for triangle and the right arrow key for rectangle. Trials were organised into blocks where a single cue (colour or shape) was repeatedly presented or in blocks where a mixture of these cues were used, and participants had to adapt flexibly to changes in cues.

Each trial began with an 800 millisecond (ms) delay with only the comparison images displayed. The condition cue was displayed between the comparison images for 800ms. A small fixation cross was displayed at the centre of the screen for 200ms to draw the participant’s attention back to the target display area from the cue display area. The target image was displayed in the centre of the screen for five seconds or until the participant responded. Up to three sets of practice blocks were used to ensure participants understood the task and could perform the task with sufficient accuracy for the output to be meaningful. Each practice block consisted of eight sequential colour trials and eight sequential shape trials. Participants advanced to the block of test trials if at least 12 of the 16 trials in one practice block were answered correctly. The accuracy score was the proportion of correct responses in the shifting block (out of 64 trials), multiplied by 5 to create a range from 0 to 5. RT data are often positively skewed and in order to adjust for this skew, a $\log_{10}$ transformation was applied to the median RT score. To further reduce skewing, the minimum RT was set to 500ms and the maximum reaction time was set to 3000ms; scores that fall outside that range were truncated (e.g., an RT of 4000ms was set equal to 3000ms and 300ms was set to 500ms). Log values were algebraically rescaled from a $\log_{500} - \log_{3000}$ range to a 0 to 5 range. The accuracy score (ranging from 0 to 5) and the adjusted RT score (ranging from 0 to 5) were summed to create the shift score. Possible scores range from 0 to 10, a higher score was indicative of better performance.

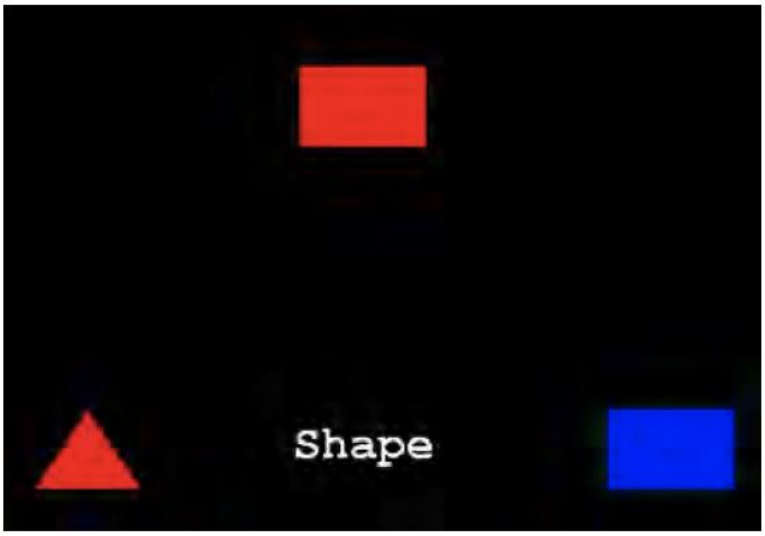


Figure 2. Example stimuli presented in the set-shifting task where participants were asked to match the target object (presented in the top and centre of the screen) to the one of the two comparison objects (presented on the bottom left and right corners of the screen) based on the condition cue (either the word colour or shape) presented on the bottom centre of the screen). Figure reproduced from Kramer et al. (2014).

Flanker Task. On each trial, participants were shown a row of five arrows on a computer screen and asked to indicate by pressing the left or right arrow button whether the centre, target arrow was pointing to the right or to the left (see Figure 3 for sample stimuli). Participants were asked to respond as quickly as possible while trying not to make any mistakes. There were two conditions, in the congruent condition all of the non-target arrows were pointing in the same direction as the target arrow and in the incongruent condition, all of the non-target arrows were pointing in the opposite direction to the target arrow. There were 24 trials of each condition and 48 trials in total. Between trials, a fixation point appeared for a random time interval between 1000ms and 3000ms and stimuli were sometimes presented above and sometimes below the fixation point. Up to three sets of practice trials were presented at the beginning, which provided feedback to participants about whether their response was correct or incorrect. Participants advanced to the test trials if they obtained at least six correct responses in a practice block.

A score combining accuracy and reaction time (RT) was calculated for the incongruent trials. The accuracy score was the proportion of correct responses (out of 24 trials), multiplied by 5 to create a score that ranged from 0 to 5. To correct for skew in RT data, the median RT was transformed in the same way as the for the set-shifting task (see above) to generate a score from 0 to 5. The accuracy score (ranging from 0 to 5) and the adjusted RT score (ranging from 0 to 5) were summed to create the flanker score. Possible scores range from 0 to 10; higher scores are indicative of better performance.

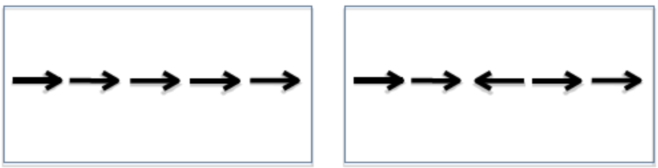


Figure 3. Examples of congruent (on the left) and incongruent (on the right) stimuli from the flanker task. Figure reproduced from Kramer et al. (2013).

Anti-saccade. The anti-saccade task is a measure of distractor inhibition. There were two distinct conditions, pro-saccades and anti-saccades. In the pro-saccade condition, participants were asked to fixate on a white circle in the centre of the screen, which appeared for 1000ms. This was followed by a blank screen for 200ms and then the white circle was displayed parallel with the fixation point on the left or right extremes of the screen for 900ms. Participants were required to shift their eye gaze to the new location of the white circle. Participants were then required to move their eye gaze back to the centre of the screen and wait for the next trial to begin. There was one pro-saccade block of 10 trials. Participants received one point for each saccade that was in the same direction in which the circled moved.

In the anti-saccade condition, a white circle again appeared in the centre of the screen for 1000ms, followed by a 200ms blank screen, and the white circle then re-appeared on the left or the right of the screen. In this condition, participants were required to shift their eye gaze in the opposite direction to where the circle had moved to. There was one block of three practice trials to ensure participants understood the task instructions followed two blocks of 10 trials. Participants received one point each time their saccade was in the opposite direction to the circle on screen, with a maximum score of 10 and 20 in the pro- and anti-saccade conditions, respectively. The pro-saccade condition was administered first to establish a pre-potent response but the score on this condition was not evaluated. Higher scores on the anti-saccade condition were taken to indicate greater difficulty with distractor inhibition. Examiners who completed the training videos provided by the assessment authors and achieved reliable scores scored these trials.

Dysexecutive errors. Executive-related difficulties can manifest as impulsive errors, failure to shift set, perseverative behaviour, and stimulus-boundedness, even when overall descriptive achievement scores on tests are unremarkable. Accordingly, the NIH-EXMAINER also generates a composite measure of dysexecutive errors committed across the administration of the whole battery. This includes the experimenter's rating of their perception of participants' behaviour with respect to stimulus boundedness, social inappropriateness, perseverative responding and motivation. These ratings were combined with the number of repetitions and rule violations in each of the semantic and phonemic fluency tasks, the false alarm responses on the CPT task, the amount of errors made on the incongruent trials relative to congruent trials of the flanker task, and the number of errors made on shift trials relative to non-shift trials in the set-shifting task.

I calculated a *working memory factor* score, according to manual instructions (Kramer et al., 2013), by combining the total dot-counting score and the d-prime (*d* `) from the spatial 1- and 2-back tasks, which are described below.

Dot counting. The dot-counting task assessed verbal working memory and was based on the counting span task (Case, Kurland, & Goldberg, 1982). Participants were presented with a mixed array of green circles, blue circles and blue squares on a computer screen and asked to count and remember the total number of blue circles. After the participant finished counting the blue circles on one screen, they repeated the total aloud to indicate that they were finished. The examiner then switched to a new screen with a novel array and participants were asked to count aloud and remember the number of blue circles on this new screen. After a set number of screens had been presented, participants were asked to recall the total number of blue circles they counted on each of the screens presented in the order in which they were presented. Participants were given one point for each correct total that a participant recalled in the correct location of a trial sequence. The task began with a practice block of three trials, to ensure participants understood the instructions, followed by six test trials where the number of display screens increased by one in each successive trial, from two to seven (see Figure 4 for an example of stimuli used in the three-screen condition). The number of correct responses was recorded and totalled across all six trials. The total possible score was 27 points. Higher scores reflected better verbal working memory.

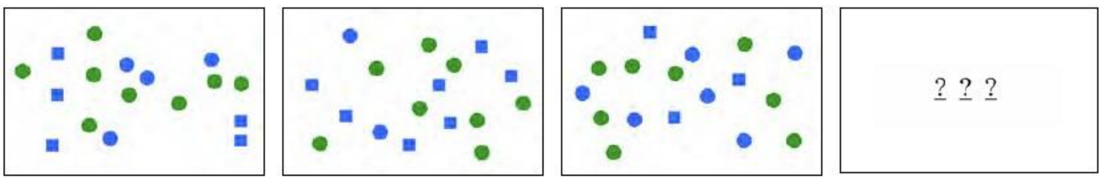


Figure 4. Example dot counting task stimuli where participants had to count aloud and remember the number of blue circles presented on each of the three screens and were then prompted to recall these totals, in the correct order. Figure reproduced from Kramer et al. (2013).

Spatial 1-back. In the spatial 1-back task, which assessed spatial working memory, the participant was asked to remember the location of a white square (the probe) that appeared in one of 15 possible locations on a computer screen for 1000ms. This was followed by a 500ms delay when the screen was entirely blank. The participant was then asked to read aloud a number that was presented in the centre of the screen for 1000ms, which ensured that the participant did not maintain fixation at the location of the previously presented probe as a memory aid. There was another 500ms delay before another white square appeared (the target). The participant was asked to indicate whether this target square was in the same location as the probe by pressing the left arrow key for 'yes' and the right arrow key for 'no'.

After the participant responded, another white square appeared which served as both a target that needed to be held in memory for the following trial and as a probe that is compared to a target displayed on the earlier trial, see Figure 5. Up to two sets of practice blocks, of 10 probes were used to ensure participants understood the task and could perform the task with sufficient accuracy for the output to be meaningful. The participant advanced to the testing block if at least seven of the 10 trials in one practice block were answered correctly. The test block consisted of 30 probes. Based on signal detection theory, the preferred measure in *n*-back tasks, such as this one, is to compute a net score that takes into account the relative proportion of hits and false alarms, called a *d* prime (d`; Swets, Tanner, & Birdsall, 1988). Possible scores ranged from -1.11 to 3.67 and higher scores are indicative of better performance. I calculated a *d* ` for the spatial 1-back task using the following formula:

$$d` = Z_{[\text{hit rate}]} - Z_{[\text{false alarm rate}]}$$

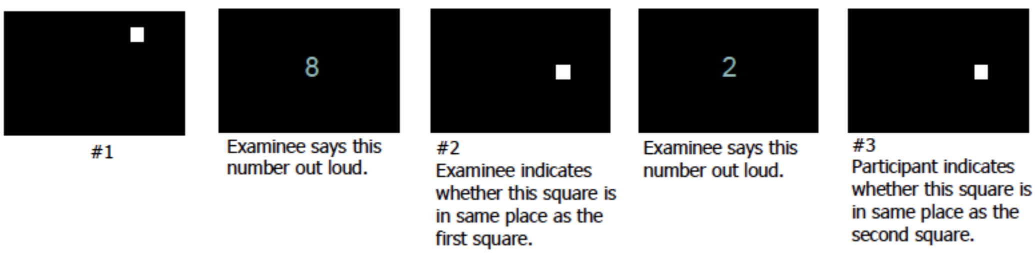


Figure 5. Example stimuli in the spatial 1-back task where participants had to remember to location of the probe on the first screen, say aloud the number presented on the subsequent screen and indicate if the target presented on the following was in the same or

a different location to the probe presented on the first screen. Figure reproduced from Kramer et al. (2013).

Spatial 2-back. For the more difficult, spatial 2-back task, the probe was compared to the target from the trial before the preceding trial (i.e., the trial 2 back from the current trial). During the 2-back task, the stimulus displayed on the preceding trial was the "intermediate" target and so no number was displayed because there was not a need to draw participants fixation away from the probe, see Figure 6. Up to three sets of practice blocks of 10 probes were used to ensure participants understood the task and could perform the task with sufficient accuracy for the output to be meaningful. The participant advanced to the testing block if at least seven of the 10 trials in one practice block were answered correctly. The test block consisted of 90 probes. A *d*` was again calculated for the 2-back task. Possible scores ranged from -1.94 to 3.88; higher scores were indicative of better performance.

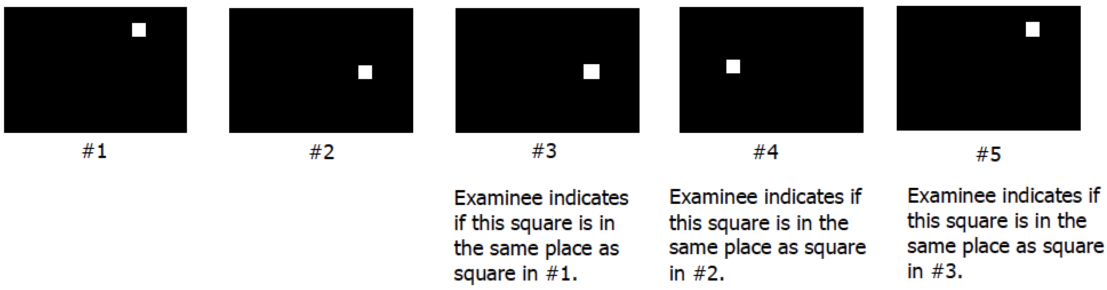


Figure 6. Example stimuli in the spatial 2-back task where participants had to remember to location of the probe on the first and second screens and had to indicate if the target presented on the third screen was in the same or a different location to the probe presented on the first screen, which had been presented two trials before. Figure reproduced from Kramer et al. (2013).

I calculated an NIH-EXAMINER *global executive factor* score, according to manual instructions (Kramer et al., 2013), by combining each of the variables that contributed to the fluency factor score, the cognitive control factor score and the working memory factor score.

There were two tasks administered (the continuous performance task [CPT] and the unstructured task) that did not contribute to the factor scores generated, because individual differences on task performance did not load clearly onto either the one- or three-factor models reported in Kramer et al. (2014). The CPT and unstructured tasks are described below.

Continuous Performance Task. In the CPT, the participant was asked to press the left arrow key, as quickly as they could, every time the target image (a five-point star) was presented but was asked not to press any key if any of the other five non-target shapes were presented (see Figure 7 for the stimuli used). Each trial began with the display of an image for a duration of 750ms. Once the image was displayed, participants could respond by clicking the left arrow key. Up to three sets of practice blocks were used to ensure participants understood the task and could perform the task with sufficient accuracy for the output to be meaningful. Each practice block consisted of 15 displays of the target image and one display of each non-target image (20 trials). Participants advanced to the test trials if at least 16 trials in one practice block were answered correctly. The testing block consisted of four sets of 25 trials (100 in total) with each set having 20 displays of the target image and one display of each non-target image. The total number of false alarm errors was recorded, the total possible score was 20. This was reverse coded so that higher scores were indicative of greater ability with respect to pre-potent response inhibition.

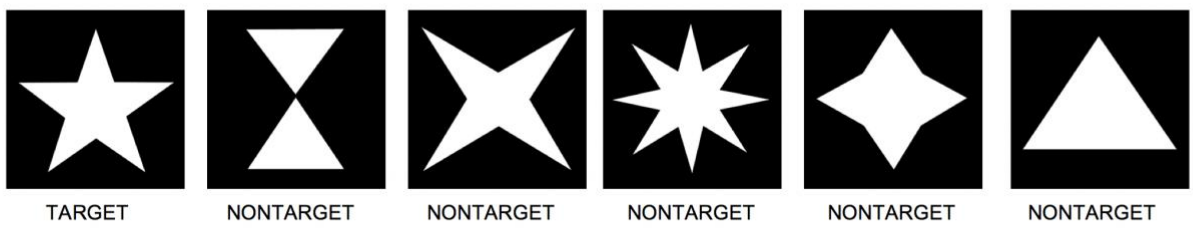


Figure 7. Example stimuli for the continuous performance task where participants had to press the left arrow key as quick as they could every time the target image appeared and without their response if any non-target image appeared. Figure reproduced from Kramer et al. (2013).

Unstructured Task. The unstructured task is conceptually similar to the Six Elements Task (Burgess et al., 1998) and designed to measure strategy generation, planning, and strategy execution. In a practice condition, participants were presented with one sheet of paper containing six puzzles and asked to complete them to ensure they understood the types of puzzles involved. In the main task, participants were provided with three test booklets, each booklet contained six pages, with four puzzles on each page (see Figure 8). In each booklet, the number of points that could be earned was identical but the number of points available on each successive page was reduced and so advancing through a booklet resulted in diminishing returns. The puzzles varied with respect to the points earned for completing them and the time they took to complete. Participants were asked to earn as many points as possible in six minutes. They were informed that they did not have to complete an entire page or an entire booklet before moving to a different puzzle and that they would only receive points if a puzzle were fully completed. A timer remained visible to participants throughout.

The total number of points earned for completing puzzles was recorded for each participant. Yet relying solely on the total number of points earned was not informative since participants who were fast could still earn a high number of points, despite making poor choices. According to Kramer et al. (2014), each puzzle was designated as high or low value based on its point value relative to the average time taken to complete it. For each participant, the number of high value and low value puzzles completed in six minutes was recorded. A possible disadvantage of relying solely on the percentage of high value puzzles is that participants who solved only one or two puzzles could potentially have inflated values for the percentage of high value puzzles. Accordingly, a weighted composite was generated that combined the total points earned with the number of high and low value puzzles completed using the formula below. Higher scores were indicative of better planning ability.

$$\left(\left(\frac{completed\ high\ value\ items}{completed\ high\ value\ items\ +\ completed\ low\ value\ items}\right) \times 100\right) \times\ \log_{10}(total\ points\ earned + 1)$$

General Procedure

Participants were seen one-on-one in a quiet research facility at UCL Institute of Education (*n* = 72) or in a quiet room at their school (*n* = 7). Completing the battery of tasks took approximately one hour and, where needed, participants were seen on multiple occasions to complete testing. Participants completed the NIH-EXAMINER battery and the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011). A parent or caregiver for each participant completed a general background questionnaire and the Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012).

The above procedure had ethical approval through the postgraduate student approval procedure at UCL Institute of Education. The ADOS-2 was not administered to all autistic participants because it was not an included measure in the planned protocol for this study due to time limitations. Data are reported here in cases where ADOS-2 data were available because participants took part in other research studies and consented for their data to be used in this way. Of the 19 participants who had ADOS-2 data available, the ADOS-2 was administered to participants during one of the visits to the research lab that they undertook to participate in this study (*n* = 11) because they were participating in another study concurrently or it was administered during a visit for a previous study (*n* = 8) when they consented for their data to be re-used. Participants completed all the testing they took part in (including a number of tasks detailed in other chapters) either in one visit (*n* = 50), two visits (*n* = 26) or in three visits (*n* = 3).

Data availability

De-identified data collected for this study are perpetually available in a publicly accessible repository (https://osf.io/ngmpb/).

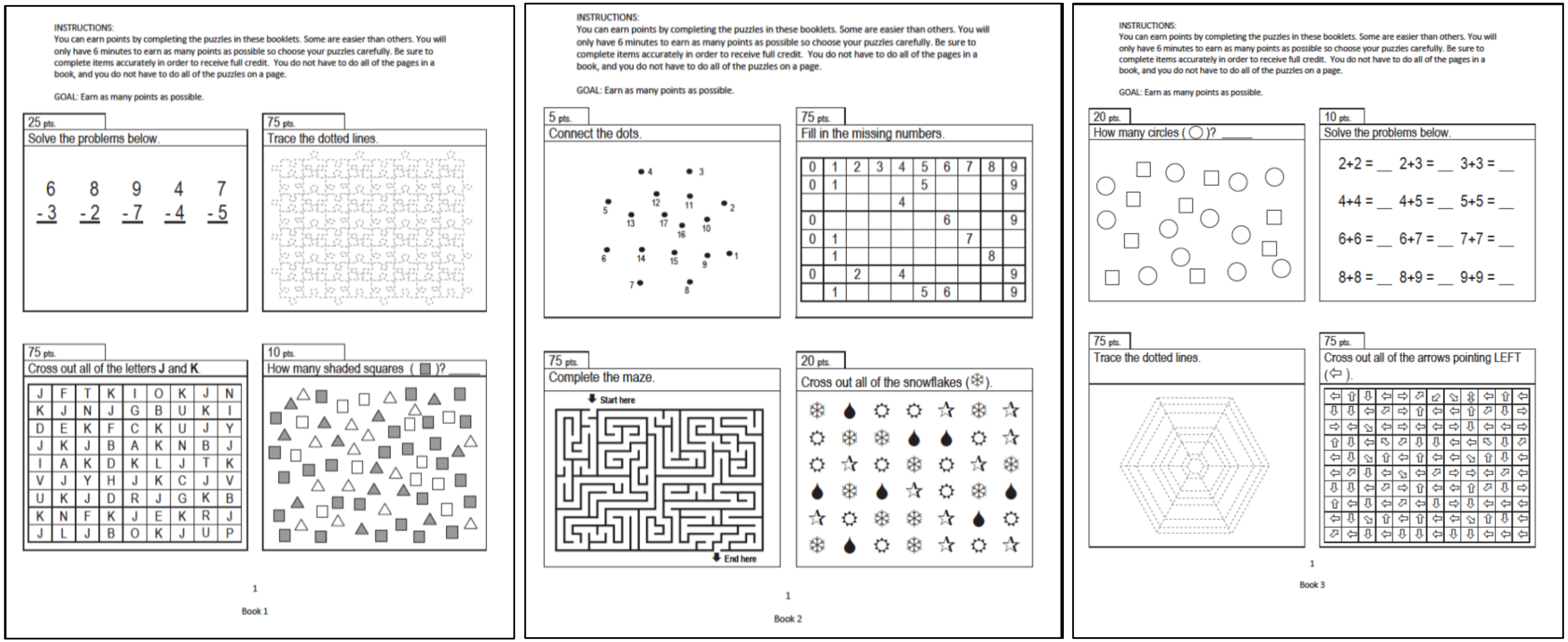


Figure 8. Page one of each of the three books of puzzles used for the unstructured task. Figure reproduced from version A of the NIH-EXAMINER battery testing forms (Kramer et al., 2013).

Results

Data screening

To begin, I assessed the distribution of each dependent variable for normality using the Shapiro-Wilk Test. Performance scores on the spatial 1-back, dot counting, and unstructured tasks were normally distributed, and all of the factor scores were normally distributed. Performance scores on the following tasks deviated from normality: phonemic fluency, semantic fluency, set-shifting, flanker, anti-saccade, dysexecutive errors, CPT the spatial 2-back. Tukey's Ladder of Powers was performed on each of the variables that were not normally distributed using the *rcompanion* package in R (Mangiafico, 2016) and successfully transformed them to be normally distributed, with the following exceptions: the spatial 2-back, the anti-saccade and the CPT, where non-parametric statistics were used to compare groups. A number of participants failed to progress beyond the practice condition in the spatial 2-back (n = 1; non-autistic) and in the flanker tasks (n = 2; both autistic).

Analysis plan

To test my first aim, that autistic participants would have more difficulties on each of the individual tasks in the NIH-EXAMINER battery, I conducted one-tailed independent sample t-tests, or Mann-Whitney U tests for non-normally distributed variables, on the dependent variables specified in the manual, for each task. The use of one-tailed hypothesis testing here reflects the directional hypothesis being tested; that is, that the non-autistic sample should have fewer difficulties than the autistic sample on each task within the NIH-EXAMINER. The extant literature has repeatedly shown autistic participants underperform relative to their non-autistic peers on assessment such as these and so assessing this directional hypothesis, rather than the non-directional alternative (that the groups would simply differ in their performance), increases the statistical power to reject the null hypothesis (Ruxton & Neuha, 2010). To test my second aim, that autistic participants would have lower scores than non-autistic participants on each of the factor scores from the NIH-EXAMINER (global executive composite, fluency factor, cognitive control factor, and working memory factor), I conducted one-tailed independent sample *t*-tests. Analyses were conducted in R (version 3.4.3; R Core Team, 2017) and RStudio (version 1.1.4; RStudio Team, 2015).

Group differences

Table 4 and Table 5 show the descriptive statistics for the dependent variables from individual tasks for the generated factor scores, respectively (non-transformed scores are presented here for clarity but transformed scores were used in group comparisons). Similarly, Figure 9 and Figure 10 show performance on individual tasks and factor scores, respectively (*z*-scores of raw scores and, where appropriate, of transformed scores are used here, so performance can be plotted on a single figure).

Table 4.
Descriptive statistics for the dependent variables specified for each task in the NIH-EXAMINER battery

	Non-autistic (n = 42)	Autistic (n = 37)
	M (SD) Range	M (SD) Range
Fluency tasks		
Phonemic fluency	22.31 (7.75) 8 - 48	20.22 (6.25) 9 - 36
Semantic fluency	30.40 (8.98) 16 - 61	26.97 (7.70) 8 - 46
Cognitive control tasks		
Set-shifting	8.18 (0.85) 5.42 - 9.39	7.78 (0.92) 5.83 - 9.62
Flanker	8.96 (0.53) 7.13 - 9.78	8.80 (0.59)[a] 7.17 - 9.85
Anti-saccade	31.43 (6.27) 15 - 40	30.65 (5.26) 19 - 40
Dysexecutive errors	10.39 (6.43) 1 - 30	11.89 (6.84) 3 - 30
Working memory tasks		
Dot counting	18 (3.97) 10 - 26	14.68 (4.41) 2 - 22
Spatial 1-back	2.28 (0.56) 1.02 - 3.67	2.15 (0.83) 0.14 - 3.67
Spatial 2-back	1.03 (0.74)[b] -0.21 - 3	0.96 (0.68) -0.23 - 2.79
Tasks not contributing to factor scores		
Continuous performance[c]	16.19 (3.89) 3 - 20	16.43 (3.78) 5 - 20
Unstructured task	162.55 (29.80) 91.63 – 234.47	143.07 (37.22) 56.44 – 210.98

Note. Phonemic fluency = total correct responses across both trials of the task, Semantic fluency = total correct responses across both trials of the task, Set-shifting = total weighted accuracy and reaction time score, Anti-saccade = total number of trails in the anti-saccade condition, Dysexecutive errors = combined score of errors committed across the administration of the NIH-EXAMINER battery, Dot counting = total correctly recalled screens, Spatial 1-back = *d* ` on the 1-back condition of the *n*-back task. Spatial 2-back = *d* ` on the 2-back condition of the *n*-back task. Continuous performance = the total number of false alarm errors recorded. Unstructured task = Weighted composite score from the unstructured planning task which combined points earned with the relative number of the high-value and low-value puzzles completed.
[a] *n* = 35. [b] *n* = 41. [c] = reverse coded scores.

Autistic participants had significantly lower scores than the non-autistic participants on the semantic fluency task, $t(77) = 1.84, p = .035, d = 0.24$. There were no significant group differences on the phonemic fluency task, $t(77) = 1.31, p = .097, d = 0.30$, or on the fluency factor score $t(77) = 1.38, p = .087, d = 0.31$, however.

Autistic participants had significantly lower scores than the non-autistic participants on the set-shifting task, $t(77) = 2.07, p = .021, d = 0.47$. There were no group differences on the flanker task, $t(75) = 1.29, p = .100, d = 0.30$, the anti-saccade task, $U = 864.00, p =$

.197, $d = 0.18$ or on the cognitive control factor score, $t(77) = 1.64$, $p = .052$, $d = 0.37$, however. The comparison between groups on the cognitive control factor score approached significance and so it remains possible that a study with larger samples may detect group differences on this variable, albeit a difference of a smaller magnitude than has previously been suggested on measures of cognitive control in autism studies.

Autistic participants had significantly lower scores than the non-autistic participants on the dot-counting task, $t(77) = 3.53$, $p < .001$, $d = 0.80$. There were no significant group differences in the spatial 1-back task, $t(77) = 0.86$, $p = .197$, $d = 0.19$, or the spatial 2-back task, $U = 787.00$, $p = .390$, $d = 0.10$ or on the working memory factor score, $t(77) = 1.45$, $p = .075$, $d = 0.33$, however.

With respect to the tasks that did not contribute to the generation of factor scores, autistic participants had significantly lower scores than the non-autistic participants on the unstructured task, $t(77) = 2.58$, $p = .006$, $d = 0.58$, and there were no significant differences in performance on the CPT, $U = 761$, $p = .565$, $d = 0.06$.

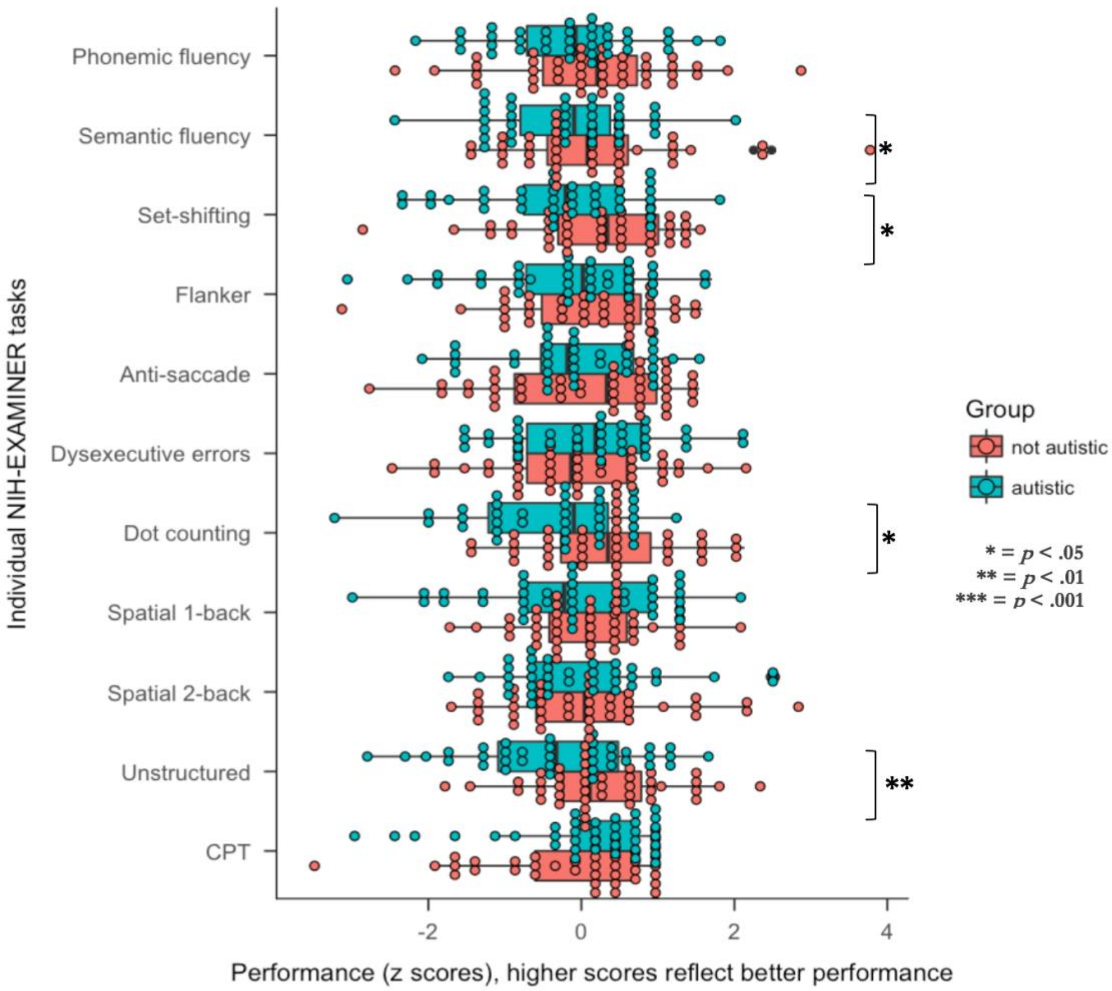


Figure 9. Performance on individual NIH-Examiner tasks, by group. CPT = continuous performance task. Performance scores on the CPT were reverse coded, such that higher scores on every measure reflected better performance.

Table 5.
Descriptive statistics for the NIH-EXAMINER factor scores for the one-factor model (global executive factor) and the three-factor model (fluency factor, cognitive control factor and working memory factor)

	Not autistic (n = 42)	Autistic (n = 37)
	M (SD) Range	M (SD) Range
Global executive factor	0.43 (0.54) -0.66 - 2.09	0.16 (0.50) -0.76 - 1.08
Fluency factor	0.20 (0.69) -1.31 - 2.34	0.00 0.59 -0.95 - 1.37
Cognitive control factor	0.73 (0.63) -0.99 - 2.05	0.50 (0.61) -0.54 - 1.89
Working memory factor	0.30 (0.61) -0.73 - 1.91	0.09 (0.67) -1.59 - 1.82

With respect to the global executive composite score, autistic participants had significantly lower scores than the non-autistic participants, $t(77) = 2.28$, $p = .013$, $d = 0.51$. That is, autistic adolescents had EF difficulties compared with their non-autistic peers on a robust measure of EF that was derived from performance across tasks on a wide-ranging battery.

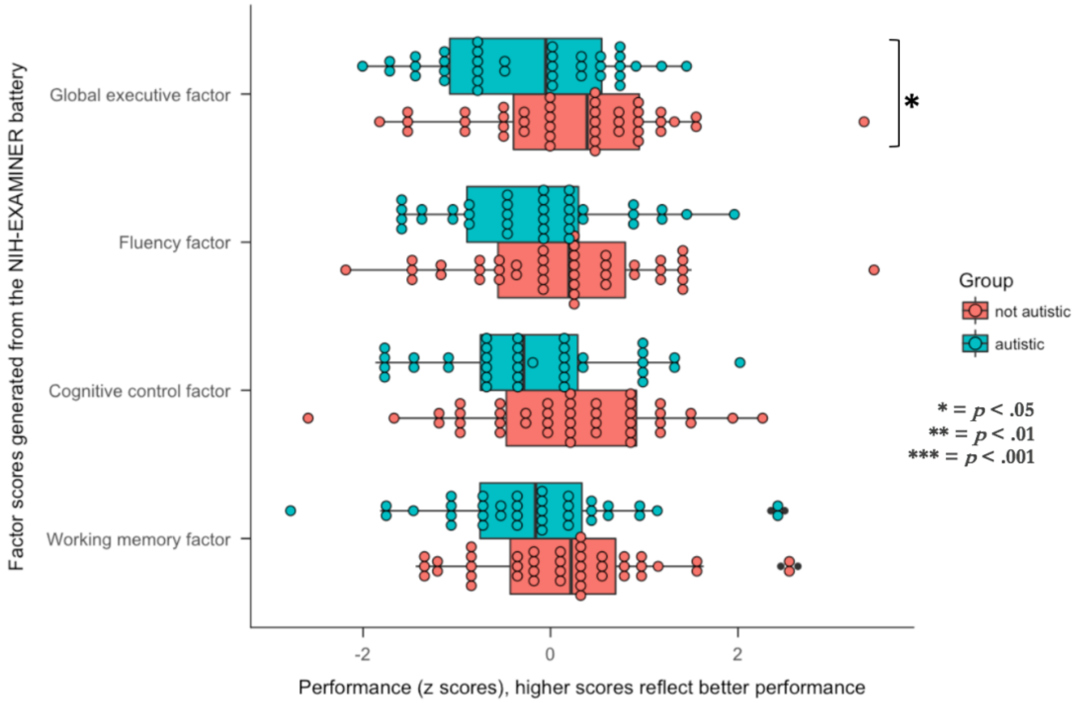


Figure 10. Performance on the factor scores generated by the NIH-Examiner battery, by group.

Individual differences

The final aim of the present study was to test whether there were relationships between performance on the NIH-EXAMINER and on other, non-executive variables. Figure 11 shows the relationships between, the global executive composite and: (A) verbal IQ, (B) performance IQ, (C) age and (D) autistic traits, as indexed by the SRS-2. I conducted a series of Pearson product-moment correlations for each of the above. Age was associated with the global executive factor for non-autistic participants, $r(40) = .48$, $p = .001$. An r-to-z transformation showed that these relationships were not significantly different, $z = -0.05$, $p = .960$. There was a significant association between verbal IQ and performance, $r(35) = .49$, $p = .002$, and between non-verbal IQ and performance, $r(35) =$

.54, $p < .001$, for autistic participants only. None of the other tested correlations were significant (all ps > .440).

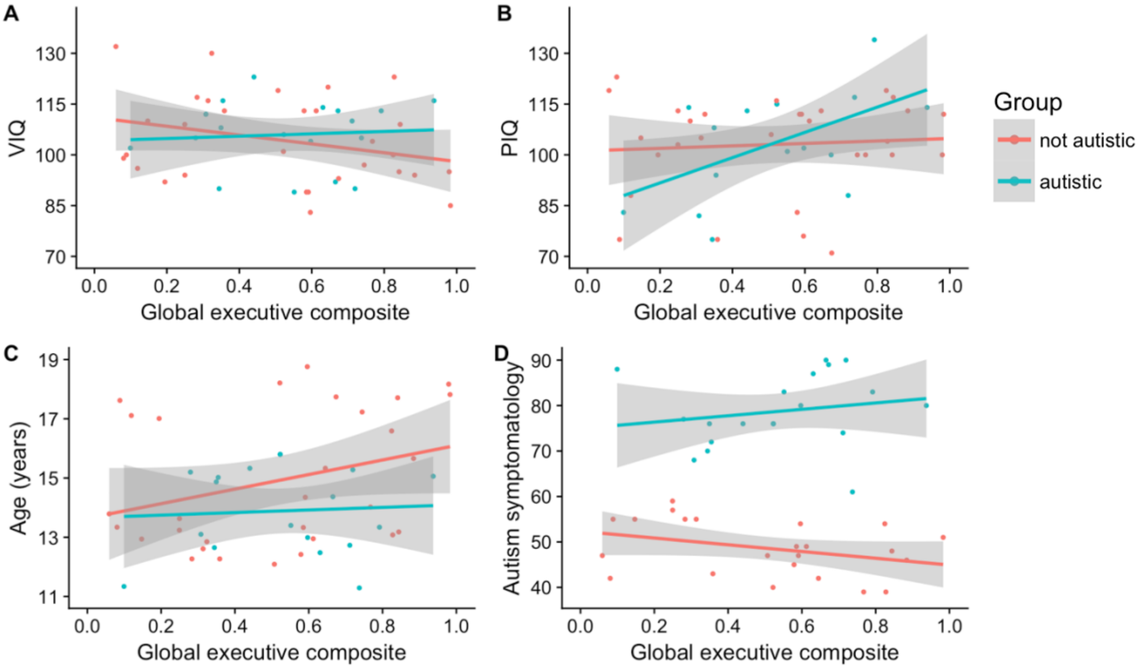


Figure 11. Relationship between performance on the global executive composite score from the NIH-EXAMINER battery and (A) VIQ = verbal comprehension index on the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011), (B) PIQ = perceptual reasoning index on the WASI-II, (C) age (in years) and (D) autism symptomatology, indexed by t-scores from the Social Responsiveness Scale, second edition (Constantino & Gruber, 2012).

Discussion

This was the first study to compare autistic and non-autistic participants on the NIH-EXAMINER battery and showed that autistic adolescents have moderate EF difficulties ($d = 0.51$) compared to their non-autistic peers on the global executive factor, which supported my hypothesis. That is, the NIH-EXAMINER battery was sensitive to the moderate, global EF difficulties previously demonstrated among autistic participants (Demetriou et al., 2018; Hill, 2004a; E. C. L. Lai et al., 2017), thus providing a justification for using this standardized battery in autism research. Future investigations can therefore use this open-source and completely standardised battery to scale up the assessment of the EF of autistic people to include large, diverse samples, to conduct like-for-like replications and to re-assess the accuracy of the conclusions previously drawn in the literature in an iterative manner, as more data is collected, shared and compiled.

Along with the global EF difficulties evident in autistic people, claims have been made for specific profiles of executive difficulties in autism, such as that autistic people have particular challenges with flexibility and planning (Hill, 2004b; Ozonoff & Jensen, 1999). More recently, some meta-analyses compared directly the results of assessments of each of the EF subdomains and concluded that, rather than a specific profile of difficulties, autistic people experience EF difficulties that apply evenly across individual domains and have an overall, rather than a fractionated, profile of EF difficulties (Demetriou et al., 2018; E. C. L. Lai et al., 2017). The supposition that followed was that autism researchers need not worry about which EF assessments to employ, as EF difficulties would be equally likely to be measurable on any EF task. The results of the current study did not support this supposition because there were no significant group differences on any of the factor scores representing individual subcomponents of EF (fluency, cognitive control and working memory). This was inconsistent with previous investigations that found autistic people had difficulties on measures of fluency (Carmo, Duarte, Pinho, Marques, & Filipe, 2015), cognitive control (Leung & Zakzanis, 2014) and working memory (Wang et al., 2017). This failed to support my hypotheses that autistic participants would show difficulties on every EF subdomain and his ramifications for how we select measures to identify which people struggle with EF abilities.

The use of latent, and not manifest EF variables might be one possible explanation for the discrepancy between the lack of group differences on the EF subdomain factors in the present analysis and the differences on tasks ranging across all EF subdomains that have been reported, albeit somewhat inconsistently, in the literature. I assessed this possibility by comparing groups on the dependent variables from each EF task in the battery (i.e., manifest EF variables). My hypotheses that autistic people would perform more poorly on each task was only partially supported. Autistic participants had more difficulties with the semantic fluency, set-shifting, dot counting and unstructured tasks. These findings are inconsistent with the lack of group differences on the fluency, cognitive control and working memory factors. Differences on these individual tasks are nevertheless consistent with previous investigations that have shown autistic people to have difficulties relative to non-autistic people on measures of semantic fluency (e.g., Spek, Schatorjé, Scholte, & van Berckelaer-Onnes, 2009), set-shifting ability (e.g., Tsuchiya, Oki, Yahara, & Fujieda, 2005), verbal working memory (e.g., Gabig, 2008) and planning (e.g., Zinke et al., 2010). The lack of group differences on the remaining tasks was inconsistent with findings that suggest, on balance, autistic people have difficulties on each of the EF subdomains represented by these tasks.

An important implication of these discrepant findings is that, had I only conducted individual tasks to measure semantic fluency, set-shifting and verbal working memory and conflated variance on manifest variables, such as these, with the latent constructs in which I am interested, I may have mistakenly assumed that this particular sample of autistic people struggle on these EF constructs. Unravelling the extent to which systematic variance in general, rather than specific EF abilities, and systematic variance in non-executive abilities as well as the extent to which non-systematic variance, or measurement error, is driving the discrepant findings is a priority before we can meaningfully assess the correlates and consequences of EF. There are four possible explanations for these discrepant findings, presented below.

First, the tasks grouped by EF subcomponent might not be reliably tapping the same construct. For example, one relevant claim is that *n*-back tasks are not a reliable metric of spatial working memory (Kane, Conway, Miura, & Colflesh, 2007), which, if true, means that the shared variance between the dot-counting task and the *n*-back task might not reflect working memory ability *per se* and might also reflect non-executive sources of shared variance, or simply measurement error. The non-executive features that differ between tasks, such as cognitive load, language demands, and task difficulty do not only impact EF performance but can also impact a participant's motivation and affective response to that task, exacerbating this issue (Poldrack & Yarkoni, 2016). Deriving factor scores across tasks where task performance is differentially sensitive to the underlying construct of interest might introduce measurement error into the factor scores rather than counteracting the task impurity problem they were derived to address.

Second, perhaps the executive difficulties reported upon in the literature diminish or disappear in adolescence. This could be because (i) autistic people learn the executive skills that they so markedly struggled with in childhood by the time they reach adolescence or (ii) all adolescents experience a drop in EF ability due to the neural re-organisation that takes place with the onset of puberty and common tasks become increasingly insensitive to difficulties through development. Support for these possibilities exists, previous reports have found executive difficulties to decrease with advancing age (Geurts et al., 2014) or to disappear altogether in one analysis for measures of cognitive control and working memory (Demetriou et al., 2018), consistent with the present analysis. Although, age does not always appear to moderate the effect sizes found, for example, in planning (Olde Dubbelink & Geurts, 2017) or working memory (Wang et al., 2017).

Third, autistic people might struggle with unstructured tasks rather than EF in general. According to White et al. (2009), an EF task is unstructured when it has been explicitly designed such that there are a number of ways of doing the task or is characterised explicitly to be open-ended as acknowledged clearly by the inventor of that task. Three tasks from the NIH-EXAMINER battery fit this definition, the phonemic and semantic fluency tasks and the unstructured task. One important finding from the current study was that autistic participants had more difficulty than non-autistic participants did on the semantic fluency task and on the unstructured task, but performance did not differ between groups on the phonemic fluency task. Autistic people struggling on two of the three less-structured tasks within this battery provides support for previous suggestions that tasks involving open-ended responses (S. J. White, 2013; S. J. White et al., 2009) and multifactorial tasks that tap planning and multitasking abilities (Hill & Bird, 2006; Mackinlay, Charman, & Karmiloff-Smith, 2006) are especially difficult for autistic participants. It therefore remains possible that specific features of assessing EF ability in the necessarily constrained environs of the laboratory mean that the NIH-EXAMINER battery and similar neuropsychological tasks might be providing much of the scaffolding – the cueing, attention orienting, and task monitoring – needed for successful task completion. The autistic participants in the current study had particular difficulties on the tasks that required them to generate a strategy, to continuously appraise and modify their behaviour and to sustain their attention without external support from the experimenter. If these are the executive abilities with which autistic people struggle the most and reliance on these abilities is low in many EF assessments, then we may be underestimating the difficulties autistic people have with executively demanding tasks beyond the laboratory.

The fourth, and final potential reason for the discrepant findings presented here is that group differences exist but the effect sizes are smaller than many studies have previously reported. It is noteworthy that the effect size estimates for several of the tasks was small but nonetheless sizeable, including for the fluency factor (d = 0.31), the cognitive control factor (d = 0.37) and on the working memory factor (d = 0.33). Based on an alpha of .05 and power of .80, sufficient statistical power to detect reliably group differences between autistic and non-autistic participants would require samples of 130, 92 or 115 participants per group, respectively. Recruiting this number of participants for the current study was not feasible with the resources and time available to me. Nevertheless, depositing the data collected from this study into a public repository will allow other researchers to add to these data, to assess samples that are increasingly large and diverse, and to test more definitively whether group differences of these magnitudes are reliable and valid.

None of the participants in the current study had an ID and the claims made herein about the EF profiles of autistic people cannot therefore be generalised to autistic people who do have an ID. Historically, EF research with autistic people aimed to uncover which individuals had EF difficulties but, more recently, interest has been growing in understanding which people can compensate for some of their cognitive difficulties through adaptation served by other cognitive skills at their disposal. IQ and EF are two such cognitive constructs which have been implicated in the ability to compensate for difficulties experience by autistic people (Livingston & Happé, 2017). Studying the EF abilities of a sample such as those who participated in the current study, including only those without an ID presents the possibility that some of the measured variability in EF task performance may represent some participant's ability to compensate for their EF difficulties by a process of adaptation that was made possible because of their high general cognitive ability or IQ. Future experimental designs should further decouple the role of IQ and EF in order to better understand their relative contributions to outcomes among autistic young people.

The advantage of using a fully manualised EF battery is that the analysis decisions needed to calculate the variables representing both individual task performance and overall EF ability were predetermined and, as such, were impervious to the large number of researcher degrees of freedom that could have been employed to impact the likelihood of producing group differences. That said, it remains unclear whether those analysis decisions outlined within the manual (e.g., summing RTs and accuracy scores) are optimal for producing group differences when comparing autistic and non-autistic participants. Future exploratory research should therefore assess the sensitivity of each of possible variable that could be generated from these tasks when comparing autistic and non-autistic participants.

In conclusion, I have established that autistic adolescents have moderate difficulties with EF ability in general but the effect sizes for each of the factor scores relating to EF subdomains were smaller in magnitude and, in many cases, did not yield significant group differences. The different pattern of performance observed on tasks that were structured and on tasks that were less structured leads to the suggestion that many classic experimental and psychometric measures of EF might not be tapping the specific difficulties that autistic people experience. Chapter 3 will test this possibility directly by assessing a subsample of those included in the current Chapter on a more open-ended, less-structured task – one that is executively demanding but more closely resembles the demands of an everyday activity, namely making tea.

Chapter 3

Comparing the executive function ability of autistic and non-autistic adolescents with a representatively-designed 'tea-making' task

Introduction

Chapter 2 showed that autistic adolescents demonstrated overall – but not specific – difficulties on the NIH-EXAMINER battery. It remains unclear, however, whether the NIH-EXAMINER battery was sufficiently sensitive to detect the sorts of difficulties young people have in their day-to-day lives. It remains possible that performance on these tasks does not necessarily relate well to everyday behavioural difficulties. Several factors suggest this possibility. First, group differences occurred on two out of the three tasks that were ill-structured and on two out of the seven tasks where responses were more structured. This may suggest the highly structured nature of most of the executive function (EF) assessments used in Chapter 2, and in the literature more broadly, might underestimate the difficulties encountered by autistic people when they need to apply EF in open-ended, ill-structured and relativistic ways, as is required in everyday life. Second and contrastingly, it remains possible that the difficulties that were apparent on some NIH-EXAMINER tasks and on the executive composite factor are artefacts of standard EF measurement paradigms that disproportionately affect autistic people. For example, these tasks are administered by an examiner which results in many of these tasks placing sociocognitive demands on participants and they may also induce social anxiety in participants. Third, the magnitude of the effects measured might not be sufficient to have a perceptible effect on people's lives. That is, the ecological validity of standard paradigms of EF assessment have not been sufficiently tested in autistic samples to assume they are valid correlates of real-world behaviour.

This chapter describes my second empirical study, which aims to address these issues directly by testing whether a sample of autistic adolescents who, on average, have overall difficulties on the NIH-EXAMINER battery also show difficulties when their EF is assessed in a context that more closely resembles the demands of daily life. I begin by re-introducing the reader to key concepts related to ecological validity before outlining the limited literature on naturalistic EF assessments in autism.

Key concepts in ecological validity research

As outlined in Chapter 1, *ecological validity* is used here to refer to the degree to which results obtained under controlled experimental conditions can be inferred to relate to results obtained in uncontrolled, naturalistic environments (Franzen & Arnett, 1997; Spooner & Pachana, 2006; Tupper & Cicerone, 1990). The degree to which a task has ecological validity is dependent on the *representativeness* or *verisimilitude* of the task (the extent to which the form and context of a laboratory task correspond to a situation encountered outside of the controlled, experimental setting) and the *generalisability* or *veridicality* of task performance (the extent to which difficulty performing on a laboratory task is predictive of having difficulties outside of the laboratory) Burgess et al., 2006; Franzen & Wilhelm, 1996; Kvavilashvili & Ellis, 2004. This Chapter will focus on representativeness/verisimilitude and generalisability/veridicality will be revisited in Chapter 4.

Many EF tasks are intentionally content-free and aim to assess cognitive ability at an abstract level and hence, involve stimuli such as geometric shapes, letters, numbers and colours. These stimuli are selected to avoid confounding performance with individuals' specific interests and motivations. For instance, tasks typically involve responding to an abstract geometric shape rather than, for example, a picture of a car, to avoid confounding performance with differences in the motivation elicited by cars. This is problematic because people never employ cognitive faculties in content-free contexts outside of the laboratory. Hence, one threat to ecological validity of a conclusion that has been levied at EF tasks is that they lack *ethological validity* (R. C. K. Chan et al., 2008). That is, the human brain is evolutionarily adapted to produce particular behaviours within particular environments and so assessments of cognition in response to abstract stimuli may not be that informative about how cognitive mechanisms function in response to the feelings and intentions somebody experiences in the environments in which a cognitive skill is deployed in the real world.

Representatively-designed EF tasks with autistic samples

One approach taken to increase the representativeness of EF assessments has been to replace the intentionally meaningless, content-free stimuli used in classic paradigms with stimuli that hold meaning to participants, for example, pictures of faces. One such attempt involved assessing autistic and non-autistic participants on a measure of cognitive flexibility where instead of flexibly adapting between response sets based on, for example, shapes and colours, they were instead required to switch their pattern of responding based on the gender and emotion shown in photographs (de Vries & Geurts, 2012). Another study, employed the classic *n*-back paradigm and used photographs of faces as the stimuli (Koshino et al., 2008). The authors found no group differences in either study but claimed these tasks might have greater ecological validity than their traditional variants. The claim that increasing the ethological validity of the stimuli used increases the generalisability of the conclusions drawn is problematic, because the authors have not specified the conditions toward which the generalisation is intended, nor have they specified how those conditions are represented in the experimental conditions. The findings presented, for example in the first of these studies, likely generalise to the ability to flexibly switch a selected response choice, at somebody else's request, based on a person's gender or based on one of two extreme emotions they are displaying, a task rarely, if ever, required in life.

One widely-used battery for assessing EF in controlled laboratory settings that addresses some concerns over ecological validity is the Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996). The BADS involves six subtests that assess flexibility, planning, problem solving and estimation by requiring participants to complete tasks that are more easily relatable to everyday goals, such as finding misplaced keys or planning a route that includes multiple stops. Autistic adults, who showed no measurable difficulties on some standard EF tasks, had more difficulties on some subscales of the BADS (Hill & Bird, 2006). While searching for lost keys or planning a route around a zoo are comparatively more ethologically valid than, for example, the Tower of London task, these BADS assessments nevertheless lack verisimilitude. The form of assessment involved drawing a proposed search strategy on a piece of paper rather than completing an actual search in a real field complete with all the sensory, motor and affective variables that could contribute to task performance. Similarly, the context of drawing a proposed search route while sitting in a quiet room is not necessarily a reliable predictor of one's propensity to be systematic when faced with the emotional, social and other situational factors surrounding losing one's keys.

On a children's version of the task (BADS-C), autistic children (aged 7 – 12 years), like autistic adults (Hill & Bird, 2006), struggled more than their non-autistic counterparts on the six parts task (a modification of the six elements task), the Zoo map task and on the key search task (S. J. White et al., 2009). The authors make a distinction between the above-mentioned tasks, which they described as being open-ended or ill structured and several standard or 'constrained' EF tasks that autistic participants had fewer difficulties completing. Compared to real-world goal pursuit, these tasks are still comparatively highly structured and so if autistic people's EF difficulties are more apparent in open-ended, ill-structured situations, then standard EF tasks likely underestimate their difficulties. Findings with the BADS are, however, somewhat mixed. For instance, autistic adolescents have been shown to be comparable to non-autistic samples on most (Harris et al., 2008) or all subtests (Rajendran, Mitchell, & Rickards, 2005).

Another laboratory-based assessment used to compare 14 autistic and 16 non-autistic cognitively able children and adolescents that emulates some demands of real-world multitasking is the Battersea multi-tasking paradigm (Mackinlay et al., 2006). Participants in this study were required to complete three simple interleaved tasks (bead sorting, counter sorting, and caterpillar colouring) in three minutes. Participants had to follow four rules such that effective performance required frequent switching between tasks, rendering sequential performance ineffective. The authors generated six dependent variables, which allowed a granular analysis of the source of potential multitasking difficulties. Overall, autistic children struggled more than non-autistic children did with multitasking, planning and switching between subtasks. While this assessment is representative of some demands of real-world multitasking (e.g., formulating a plan, shifting between sub goals, and monitoring performance relative to time), it does not represent other demands well, for example, sorting beads or colouring in are potentially not very motivating activities for the age-group assessed. Besides, performing the task in a quiet, structured environment, overlooked by an examiner likely aids performance.

A modified version of the Predicaments Test, which involves participants watching videos of 'real-life-type' social scenarios, recount pertinent facts and generate solutions to the predicaments involved, has been used with cognitively able autistic children (Channon & Crawford, 1999) and adolescents (Channon, Charman, Heap, Crawford, & Rios, 2001), who had more planning difficulties, attempted fewer tasks, switched inflexibly and broke more rules. Yet, our ability to interpret this task is hampered for three important reasons. First, the task is inherently social, and we cannot disentangle the role of executive and socio-cognitive abilities in task performance (Kenworthy et al., 2008). Second, the task involved videos of actors reading scripts, which is assumed to have greater representativeness than say, for example, a written vignette of the same scenario, but the validity of this assumption is untested. Finally, a video of an artificially created scenario is still relatively impoverished, compared to the rich, complex and nuanced ways in which real-life predicaments occur.

The Challenge task is an unpublished semi-structured observation measure where participants engage in several scenarios tapping planning and flexibility and their behaviours are rated from 0 to 3, similar to the way in which the ADOS-2 is used to elicit social difficulties. For example, children are invited to mould a figure out of clay and then asked to swap it with the examiner before it has been completed, the flexibility of their behaviour is then rated. One published account of this task demonstrated that autistic children who underwent an intervention called Unstuck and On Target, significantly improved with respect to their flexibility score on the Challenge Task (Kenworthy et al., 2014).

An unpublished measure of EF, called the Ecologically valid Test of Executive Dysfunction (Eco-TED), has been designed to address some of the ecological validity concerns specific to EF and to be a feasible tool for the clinical assessment of autistic children (Bristow, 2016; Pullinger, 2017). The Eco-TED includes seven subtests, which are a combination of paper-and-pen, action imitation and construction tasks that tap prospective memory, working memory, planning abilities as well as the ability to execute multiple-step goals. Autistic children made more errors, relative to their non-autistic peers, on some Eco-TED tasks, for example, the school bag task, where participants had to point to pictures of items that they would place in their school bag in preparation for a scenario that examiners read to them. This battery is early in development and holds some promise, but it is however subject to many of the limitations previously listed, such as limited ethological validity and providing much of the scaffolding needed for successful performance.

Previous investigations showed that frontal lesion patients complete fewer goals than samples without a Dysexecutive syndrome in both virtual and real instantiations of an office-based multiple errands task, leading to the suggestion that both are sensitive to executive difficulties (McGeorge et al., 2001; Rand, Rukan, Weiss, & Katz, 2009). The virtual errands task was modified to be developmentally appropriate for younger participants by substituting the office with a school setting (Rajendran et al., 2011). Autistic adolescents, assessed with this paradigm, were no different to their non-autistic counterparts in their ability to complete the individual tasks presented but they engaged in behaviour that was more rigid, completed fewer tasks and broke more rules when the tasks were interleaved and presented as a multiple errands task. It is worth noting, the social demands required by the traditional multiple errands task have been removed, increasing our confidence that difficulties in task completion are related to EF rather than socio-cognitive differences. These EF difficulties, according to the authors, are more readily generalisable to everyday life than are difficulties measured by some more widely used paradigms. There has been another virtual reality task that nevertheless showed no difference between autistic and non-autistic participants in a goal-directed treasure hunt (Fornasari et al., 2013).

The virtual multiple errands task has several strengths regarding the representativeness of the testing situation, including increased ethological validity. Completing errands for a teacher has more relevance to an adolescent than does, for example, sorting shapes. The task also measures aspects of the open-ended and relativistic demands of multitasking not tapped in other assessments, such as strategic goal setting, chaining multiple goals together and managing goal contentions. It does fall short on representativeness in some important ways, however. For instance, the sensory and proprioceptive aspect of the task, that is, the book did not feel or smell like a book, the motor co-ordination of using a joystick is not equivalent to climbing stairs and pretending to complete errands for an imagined teacher is not the same as doing it for a familiar teacher. In addition, samples in this study were matched on neither performance nor full-scale IQ and so it remains possible that performance reflects group differences in performance IQ and not EF.

The Dresden Breakfast task has been used to compare the prospective memory ability of 25 autistic and 25 non-autistic young adults, matched for age and ability, in a naturalistic setting (Altgassen, Koban, & Kliegel, 2012). Participants were required to set a table and prepare breakfast while following a set of rules. The autistic participants completed fewer tasks successfully, followed rules less closely, were less efficient in their task organization and monitored the elapsing time less closely than a non-autistic sample. Interestingly, Altgassen et al. (2012) found no group differences on the Tower of Hanoi or Go-/NoGo tasks, which purport to measure planning and inhibition ability, in the same samples. Yet, of note, there were very few relationships between this task and the standard EF measures used. The analyses presented were specific to prospective memory and so it is unclear the extent to which individual differences on overall task competence related to the standard EF tasks.

While there have been examples of studies where autistic participants have been assessed on tasks that were high in verisimilitude, and on tasks that had demonstrated some degree of veridicality, there have not yet been any studies in which an assessment has been found to be simultaneously high in veridicality, verisimilitude as well as retaining acceptable internal validity. To that end, in the current chapter, I used an EF assessment that has not yet been used to assess autistic people and has only appeared once in the peer-reviewed literature to date, to assess those with 22q11.2 Deletion Syndrome (M. Schneider et al., 2016). This assessment is detailed below.

In sum, the ecological and ethological validity of standard EF tasks have been criticised because the form and context of tasks do not represent situations beyond the assessment well, and the structured nature of EF assessment removes executive demands that are likely important (e.g., Burgess et al., 2006). Representatively designed EF assessments have addressed some concerns about the representativeness of task demands

(Altgassen et al., 2012; Rajendran et al., 2011) but often-involved small, sometimes unmatched samples, and in many cases, activities that were still, to an extent, arbitrary, virtual or abstract.

Current study

My first aim was to test whether autistic adolescents had EF difficulties when they were assessed in an environment closely resembling daily life. To address this aim, I compared autistic and non-autistic adolescents (aged 11 – 19 years), matched on age and ability, on a Modified version of the Multitasking Evaluation for Adolescents (MMEA; M. Schneider et al., 2016). Specifically, participants were asked to prepare food, hot drinks and study materials for a study session in a 20-minute period. The MMEA is representative of the form in which EF abilities are employed (e.g., it involves working with real food, moving around an actual room and not in a computerised analogue) and the context in which EFs are employed (e.g., time-limited preparation for an impending event). The MMEA required participants to execute a self-generated plan that involved the simultaneous application of, and dynamic interplay between, various executive and non-executive skill.

Successful performance relied on non-executive abilities, such as language comprehension (e.g., reading and assimilating the information in the task instructions), the ability to manage stress and anxiety (e.g., remaining calm in a pressurised environment), fine motor skills (e.g., opening the packet of ham or cheese), gross motor skills (e.g., moving around the room and coordinating actions). Successful performance also, critically, required core executive function abilities, including working memory (e.g., holding task instructions and the participant-generated plan in mind), cognitive flexibility (e.g., adjusting a plan when a change to the instructions was delivered), inhibition (e.g., inhibiting automatic execution of task elements according to their own preferences instead of the provided specifications). Successful performance also relied on higher-order executive function abilities, including problem solving (e.g., deducing that the participant should include themselves), multitasking and performance monitoring.

Given that autistic participants have greater EF difficulties (e.g., Demetriou et al., 2018; E. C. L. Lai et al., 2017), I predicted that they would show less competence, on average, than non-autistic participants on the MMEA. Specifically, I predicted they would complete fewer task goals, executed goals would be of poorer quality, they would perform fewer steps for the goals attempted and would make more executive-type errors.

The degree to which a task is motivating and meaningful for somebody affects the ethological validity of that assessment. My second aim was therefore to test whether there were group differences in participants' disposition toward this task and the metacognitive ability needed to appraise task performance. To address this aim, participants were asked to rate their disposition toward completing the task before and after participating and to appraise their own performance upon task completion. Based on the EF difficulties of autistic people (Demetriou et al., 2018; Hill, 2004b), evidence that other groups with compromised EF ability lack insight (Burgess et al., 1998) and suggestions that autistic people might lack insight into their own cognition (U. Frith & Happé, 1999; Zahavi, 2010), I predicted that autistic participants would be less accurate at task appraisal. Given the dearth of research focusing on ethological validity of EF assessment in autism, I did not have specific predictions about group differences in disposition toward the task.

My third aim was to test whether individual differences in participant attributes were related to task performance. Specifically, I aimed to test whether there were relationships between performance on the MMEA and verbal IQ, non-verbal IQ, age or autism symptomatology. I was particularly interested in assessing whether these relationships differed across groups. Given that opportunities to practice the types of activities included in this task (preparing tea and food) increase with advancing age, I predicted that increasing chronological age would be related to task performance. On the basis of previous theoretical and empirical accounts (Akbar, Loomis, & Paul, 2013; Russell-

Smith, Comerford, Maybery, & Whitehouse, 2014; Russell, 1997) and the fact the task instructions were conveyed through the medium of language, I predicted higher verbal IQ would be associated with better task performance. I had no a priori hypotheses about the remaining factors and so these analyses were exploratory.

Method

Participants

Seventy-five adolescents were assessed on the MMEA, aged between 11 and 19 years (M = 14.53, SD = 1.96), including 37 non-autistic (16 female, 21 male) and 38 autistic (12 female, 26 male) participants. Of the 79 participants who participated in the study described in Chapter 2, 12 participants were not involved in this study either because they could not be re-scheduled for assessment (n = 10) or did not to consent for this aspect of the study upon reading the task instructions (n = 2). To retain statistical power and match the groups on background variables, eight participants who did not participate in the study described in Chapter 2 were recruited to take part in this study. Inclusion and exclusion criteria were comparable with Chapter 2. Participants completed all the testing they took part in (including a number of tasks detailed in other chapters) in either one visit (n = 47), two visits (n = 25) or in three visits (n = 3). Method of recruitment, participants' co-occurring conditions, medication use, and participant ethnicities can be seen in Table 6. One of the autistic participants and seven of the non-autistic participants did not speak English as their first language.

Parents reported on participants' ethnicity. There were some noteworthy differences in the distribution of ethnicities between groups, including that almost three quarters of autistic participants were from White backgrounds compared with less than half of the non-autistic participants. A Fisher's exact test revealed that these differences were significant, p = .017 (see Table 6).

Table 6.
Demographic information about the participants assessed on the Modified Multitasking Evaluation for Adolescents, by group

	Non-autistic (*n* = 37) *n*	**Autistic** (*n* = 38) *n*
Recruited through		
Public engagement and research participation event	36	26
Community contacts	1	6
School-research partnership	0	6
Parent-reported co-occurring conditions		
Attention deficit hyperactivity disorder	0	10
Dyslexia	0	9
Developmental coordination disorder	0	7
Sensory processing disorder	0	3
Irlen Syndrome	0	1
Obsessive compulsive disorder	0	1
Complex language disorder	0	1
Depression	1	0
Medication		
ADHD	0	6
Antipsychotic	0	1
Sleep inducing	0	4
Antiepileptic	0	1
Antidepressants	1	2
Non-psychoactive	2	1
Ethnicity		
Any White background	15	27
Any Asian background	10	1
Any Black background	3	4
Any mixed background	6	4
Other ethnic group	1	1
Missing or prefer not to say	2	1

Note. Any White background = White British, White Irish or any other White background; Any Black background = Black British, Black African, Black Caribbean or any other Black background; any Asian background = Chinese, Indian, Pakistani, Bangladeshi or any other Asian background; Mixed/multiple ethnic groups = Mixed White and Asian, Mixed White and Black African, Mixed White & Black Caribbean, Any other Mixed background.

Demographic data for the final sample can be seen in Table 7. Data for one additional autistic participant were excluded because they obtained an SRS-2 *t*-score of 56, which fell below the threshold indicative of being autistic (a *t*-score equal to or above 60; Constantino & Gruber, 2012; ADOS-2 data were unavailable for this participant). There were five (all female) additional participants who were recruited to be non-autistic participants. Their *t*-scores on the SRS-2, however, were between 60 and 69, which is above the threshold indicative of being on the autism spectrum and potentially reflects clinically-significant autistic features. These participants' data were therefore excluded from all analyses. All participants were cognitively able by virtue of having verbal, nonverbal and full-scale IQ scores of 70 or above. Participants who completed assessments that were included in Chapter 2 only completed the WASI-2 once. These data were included here and in subsequent analyses, as required. An additional five autistic participants were excluded for obtaining IQ scores below this threshold. The final samples were group-matched by age, $t(73) = 0.78$, $p = .437$, $d = 0.18$, verbal IQ, $t(73) = 1.22$, $p = .227$, $d = 0.28$, nonverbal IQ, $t(73) = 0.62$, $p = .539$, $d = 0.14$ and by gender distribution, χ^2 (1, $n = 75$) = 0.65, $p = .421$, $\phi = 0.12$.

Parents/caregivers (hereafter, 'parents') reported on their socio-economic (SES) status, as in Chapter 2. There was a significant group difference on the decile rank of deprivation, $t(63) = 3.05$, $p = .003$, $d = 0.76$. Parents of autistic participants were, on average, significantly more likely to be from less deprived areas ($M = 7.30$, $SD = 2.72$) than the parents of non-autistic participants ($M = 5.09$, $SD = 3.11$). There were no significant differences between the autistic and non-autistic participants in level of parental/caregiver education, $t(68) = 0.02$, $p = .984$, $d < 0.01$. There was no difference on the SES composite variable, $t(58) = 1.39$, $p = .170$, $d = 0.36$. See Table 7 for more.

Table 7.

Descriptive statistics for the developmental and background variables for the participants assessed on the Modified Multitasking Evaluation for Adolescents, by group

	Non-autistic (n = 37) *M* (SD) Range	Autistic (n = 38) *M* (SD) Range	*p*-value
Developmental variables			
Gender (M:F)	21:16	26:12	.421
Age (in years)	14.36 (1.97)	14.71 (1.95)	.437
	12 - 18	11 - 19	
Full-scale IQ	105.05 (11.50)	103.82 14.38	.682
	77 - 130	76 - 132	
Verbal IQ	105.08 (13.14)	101.39 (13.08)	.227
	78 - 132	73 - 123	
Nonverbal IQ	103.35 (12.75)	105.66 (18.91)	.539
	75 - 123	75 - 154	
SRS-2	47.55 (6.27) [a]	77.47 (9.40)	**< .001*****
	37 - 59	56 – 90 [b]	
ADOS-2 severity score [c]	- -	5.90 (2.63)[d]	-
	-	2 – 10[e]	
SES of primary caregiver			
Years of post-16 education	5.06 (4.83)[f]	5.08 (3.46)[g]	.984
	-1 – 19[h]	0 - 16	
IMD decile	5.09 (3.11)[i]	7.3 (2.72)[j]	**0.003****
	1 - 10	1 - 10	
Composite SES score	-0.12 (0.76)[k]	0.12 (0.52)[l]	.170
	-1.29 – 1.09	-1.17 – 1.33	

Note. Full-scale IQ = Full-scale IQ, 4 subtest version, derived from the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011), where mean score is 100 and standard deviation is 15; Verbal IQ = Verbal comprehension index derived from the WASI-II (Wechsler, 2011), where mean score is 100 and standard deviation is 15; Nonverbal IQ = Perceptual reasoning index derived from the WASI-II (Wechsler, 2011); SRS-2 = *t*-score, calculated separately, by gender from the Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012); ADOS-2 = the Autism Diagnostic Observation Schedule, second edition (Lord, Rutter, et al., 2012); SES = Socioeconomic status; IMD = Index of Multiple Deprivation (T. Smith et al., 2015); Composite SES score = mean of z-scores derived from the years of post-16 education and the IMD decile rank

[a] n = 33. [b] one participant fell below the threshold indicative of an autism diagnosis (a *t*-score of 60) but were retained in the analysis because they received a score above the threshold indicative of an autism diagnosis on the ADOS-2. [c] The ADOS-2 was only administered with participants already diagnosed with autism. [d] n = 19, The ADOS-2 was not administered to all autistic participants because it was not an included measure in the planned protocol for this study due to time limitations. Data are reported here in cases where ADOS-2 data were available because participants took part in other research studies and consented for their data to be used in this way. [e] Two autistic participants received ADOS-2 severity scores below the threshold indicative of an autism diagnosis (a severity score of 2 or below) but were retained in the analysis because they received a score above the threshold indicative of an autism diagnosis on the SRS-2. [f] n = 33. [g] n = 37. [h] One parent reported leaving full time education aged 15 years and so received a score of -1 when number of years of post-16 education was calculated. [i] n = 32. [j] n = 33. [k] n = 28 [l] n = 32.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Materials

Participants were tested in a room with two tables, one square in the centre of the room and one rectangular along one wall of the room. On one table, there was a filled kettle, 10 cups, a printed set of task instructions, a countdown timer and a box containing the additional materials required to complete the task (see Table 8 for a full list and quantity of the exact items that were provided). Consistent with Schneider et al. (2016), there were a number of distractor items in the box that were edible (including stock cubes, a lemon and a jar of chocolate spread) or non-edible (including envelopes, hand cream, a box of plasters and body cream). The second table was surrounded by several chairs, which the participant could use to set up for the study session, see Figure 12 for a photograph of

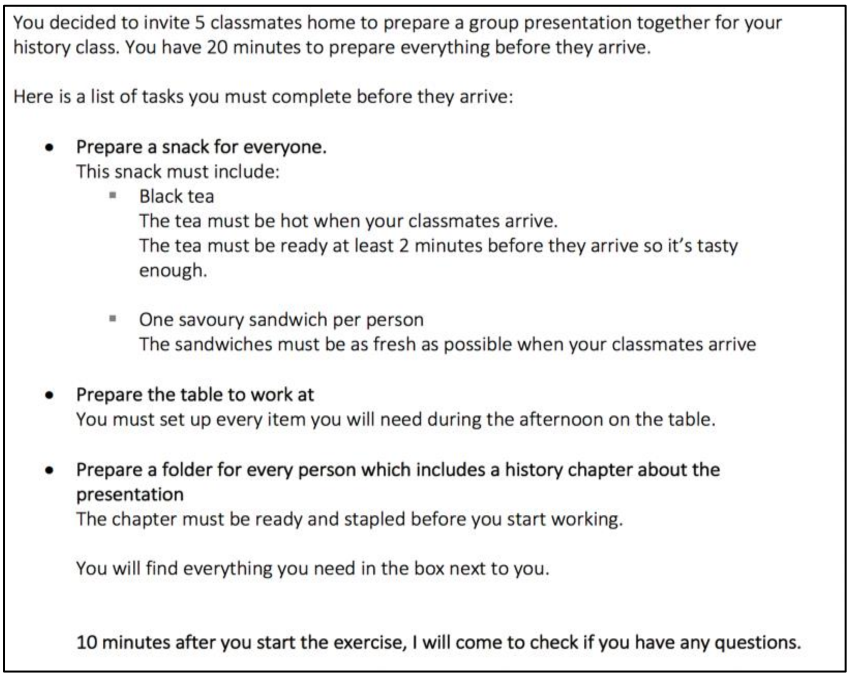
the room after two participants had completed the task.

Figure 12. Photographs of two participants' attempts at the MMEA, where they were given 20 minutes to prepare a sandwich, cup of tea and folder that contained a photocopied history chapter for a study session with five other students. The photographs show the side table (A) and central table (C) after one participant executed the task well and the side table (B) and central table (D) after one participant executed the task poorly.

Procedure

The following procedure had ethical approval through the postgraduate student ethical approval procedure at UCL Institute of Education. Participants were seen individually in a quiet research facility at UCL Institute of Education (*n* = 69) or in a quiet room in their school (*n* = 6). Participants completed a modified version of The Multitasking Evaluation for Adolescents (MMEA; M. Schneider et al., 2016) and the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011). A parent or caregiver for each participant completed a general background questionnaire and the Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012). Participants also completed a series of self-report questionnaires, the results of which will be discussed in Chapter 4. Participants completed all the testing they took part in (including a number of tasks detailed in other chapters) either one visit (n = 47), two visits (n = 25) or in three visits (n = 3).

The Modified Multitasking Evaluation for Adolescents (MMEA). Following the methodology of Schneider et al. (2016), the task began with participants receiving printed task instructions (see Figure 13). The instructions stated that the participants had to prepare for a group study session for five classmates, who would be arriving in 20 minutes. The preparation required four multi-step goals to be completed before the classmates arrived, including preparing: (i) one savoury sandwich, (ii) one cup of black tea, (iii) one folder containing a stapled photocopy of a textbook history chapter for each person, as well as (iv) set the table for the study session. Participants were not explicitly prompted to prepare materials for themselves but were expected to deduce that they would require a set of materials for themselves, as a member of the group study session (i.e., a sandwich, a cup of tea, a folder and to set a place at the table).


You decided to invite 5 classmates home to prepare a group presentation together for your history class. You have 20 minutes to prepare everything before they arrive.

Here is a list of tasks you must complete before they arrive:

- **Prepare a snack for everyone.**
  This snack must include:
  - Black tea
    The tea must be hot when your classmates arrive.
    The tea must be ready at least 2 minutes before they arrive so it's tasty enough.
  - One savoury sandwich per person
    The sandwiches must be as fresh as possible when your classmates arrive
- **Prepare the table to work at**
  You must set up every item you will need during the afternoon on the table.
- **Prepare a folder for every person which includes a history chapter about the presentation**
  The chapter must be ready and stapled before you start working.

  You will find everything you need in the box next to you.

  10 minutes after you start the exercise, I will come to check if you have any questions.

Figure 13. The printed instructions that participants were given on a sheet of paper. This sheet of paper was left visible and accessible to the participant throughout the task.

Once a participant appeared to complete one task goal in full (e.g., they finished preparing folders and moved on to preparing sandwiches), the instruction update was delivered. The instruction update involved a researcher interrupting the participant to inform them that two of the classmates could no longer attend the study session. The instruction update served to assess participants' ability to adapt flexibly to a modification in task demands after they had generated and began executing a plan. The items-to-complete were listed in a set order (tea, sandwiches, set table and folders) but the goals were not numbered. Participants were asked to ensure that sandwiches were as fresh as possible, and that the tea was hot when their classmates arrived. Participants were therefore left to decide the order in which they completed the tasks, the optimum strategy was to prepare the folders first, which necessitated reading the instructions to the end and formulating a plan before beginning any of the goals. Participants did not always attempt goals in the same order, nor did they take the same amount of time to complete their first goal and so the update in task instructions was delivered after differing amounts of time had elapsed for different participants.

Participants were told they had 20 minutes to complete the task. A countdown timer was started, and both the printed instructions and timer remained visible throughout the task. After 10 minutes had elapsed, the researcher asked the participant whether they had any questions about completing the task. If a participant asked a question at the designated question time, the researcher would respond informatively (e.g., by affirming that they should include themselves). If participants asked any questions outside of the allotted questioning time, however, the researchers always responded by saying "you should do whatever you think is correct", following Schneider et al. (2016).

Finally, there were sufficient materials present to prepare items (cups of tea, sandwiches and folders) for more people than participants were required to prepare (in this case, for up to 10 people). This was so to ensure that correct task performance relied on successfully holding task instructions in mind and appropriately modifying the instructions as required rather and that performance was not guided by cues present in the environment, such as the number of cups that were visible. See Table 8 for a full list of the items provided to participants.

To assess participants' metacognitive ability and disposition toward the task they were asked, after they read the task instructions but before they began the task, to rate on a scale ranging from 1 (not at all) to 5 (very/a lot): (i) their motivation while completing the task, (ii) how difficult the task was, (iii) how effortful they thought the task was to complete and (iv) how much they enjoyed completing the task. After participants had completed the task and the timer had been stopped, they were again asked to rate each of the above questions. Participants were also asked to rate on a scale ranging from 1 (not well at all) to 5 (very well) how they believed they performed in the task. Participants were video recorded with their prior permission. Two participants decided they would rather not be recorded, and their performance was rated as accurately as possible, immediately after they completed the task. Recordings were then coded according to the protocol detailed below.

Table 8.
List and quantity of items provided to participants completing the MMEA

Items provided	Number
Sandwich	
Full and open packet of sliced cheese	1
Tomato	2
Full and open packet of (vegetarian) ham	1
Loaf of bread	1
Container of easily spreadable butter	1
Plastic plates	10
Forks	10
Knives	10
Packet of napkins	1
Bottle of ketchup	1
Tea	
Box of teabags	1
Cups	10
Kettle (full of water and plugged in)	1
Folder	
Empty notepads	10
Plastic filing folders	10
Copies of 8 pages from history book (not collated)	10
Pens	10
Stapler	1
Distractors	
Envelopes	10
Tube of hand cream	1
Tube of body cream	1
Box of plasters	1
Lemon	1
Packet of stock cubes	1
Jar of chocolate spread	1

Changes to administration. The way in which the assessment was administered was modified significantly from Schneider et al. (2016). First, I removed one element (a toaster) as a safety precaution. Accordingly, the task could now be performed in a serial order within the 30 minutes allowed, removing the necessity to multitask. To account for this, and to maintain task difficulty, administration time was reduced to 20 minutes. Second, unlike Schneider et al. (2016), in the current study, participants were not given an opportunity to ask questions before beginning the task in order both to allow participants independently to generate a plan and to increase the importance of inhibiting the desire to ask questions until the allotted time. Third, participants were asked if they had any questions after 10 rather than after 15 minutes, this was because of the reduced administration time. Finally, two additional modifications were made to the task instructions update that was delivered part-way through the task: (a) participants were told two classmates – rather than one – could not make it in order to avoid ambiguity about whether a participant had appropriately responded to the update or omitted themselves in their preparations; and (b) the update was delivered when the first goal was completed rather than after 15 minutes had elapsed, which served to standardise the amount of plan generation and execution participants had engaged in prior to the instruction update.

Changes to the scoring protocol. The most significant change from Schneider et al. (2016) was the modifications made to the scoring criteria. The need to simultaneously deploy a variety of executive and non-executive skills was intentionally being assessed by the MMEA, but it meant that, unlike standard EF assessments, specific variables to represent either subcomponents or EF ability in general could not reasonably be calculated. Instead, Schneider et al. (2016) created a variable that reflected overall goal fulfilment. The extent to which you fulfil goals in an executively-demanding task should represent EF abilities, non-EF abilities and the ability to deploy these abilities in concert. The authors achieved this by sub-dividing each goal (sandwiches, tea, folders, setting the table) into a series of units worth one point each (e.g., buttering bread = 1 point, adding cheese = 1 point, adding tomato = 1 point, etc.) to create a total performance score out of 43. The authors also reported several general outcomes. For instance, a participant was given one point if they (i) accounted for the update in task instructions, (ii) if they included themselves or (iii) if they asked questions outside of the allotted time or they were not awarded one point if they failed to do so.

Initial pilot testing for the current study revealed that participants identified that it was not necessary to complete every step to satisfy the task instructions as they were presented. For example, the most efficient behaviour to produce based on the instruction to prepare a savoury sandwich is to make a sandwich with one filling and not to butter the bread and so penalising a participant for not including each possible step was deemed inappropriate. Consequently, rather than calculating a total performance score and reporting general items as per Schneider et al. (2016), I created four continuous outcome variables that captured different aspects of goal fulfilment and task execution. These variables were (i) the number of task goals achieved, (ii) for the attempted goals, the level of involvedness, (iii) for the attempted goals, the quality of the strategy and its execution and (iv) the number of executive errors committed. The protocol for generating each variable is detailed in turn below.

Participants were initially instructed to prepare materials for five classmates and were expected to realise the need to include themselves. This meant that correct number of materials to prepare before the instruction update was delivered, was six and after the instruction update, was four. To account for this, the dependent variables of interest, in each case below, were the proportion of possible points that could have been awarded in each category, possible scores ranged from 0 to 1, where higher scores were indicative of receiving a greater proportion of all possible points available.

Goals achieved. For each element of each goal (e.g., prepare one sandwich), participants were given a score of 0 (not completed), 1 (partially completed) or 2 (completed in full) based on the extent to which they completed the minimum necessary steps to achieve the described task. The necessary steps were operationalised in each case and every permutation of successfully completing an action was considered. For example, for making one sandwich, two points were awarded if one slice of bread was cut in half or if it was folded over or if two slices of bread were put together and at least one of the provided savoury fillings were included. Whereas, one point was awarded if an attempt was made to make a sandwich, but it was not as complete as described above, for example, the bread was buttered but the slices of bread were never placed together, or bread was laid out but never buttered. Similar criteria were applied to awarding points for preparing a cup of tea, a folder and setting the table (see Appendices

Appendix A for full scoring criteria).

Involvedness score. For each goal that was completed or partially completed, an additional variable was created to determine the level of involvedness at which the goal had been executed. That is, how many steps, whether critical to goal completion or not, were performed. One point was given for each one of a list of predefined steps that could have been correctly completed while completing each of the goals (preparing cups of tea, preparing sandwiches, preparing folders, and setting the table). For example, regarding making a sandwich, placing one slice of cheese inside two slices of bread was sufficient for the participant to receive full credit for completing the goal but they would receive only one point for including the cheese. Instead, a participant who buttered the bread and placed ham, cheese and sliced tomato inside would receive full credit for completing the goal and received one point each for including the butter, ham, cheese and tomato (see Table 9 for a full list of steps that were awarded a point for involvedness when making a sandwich). A full list of the predefined items that contributed to the involvedness score for each of the goals can be seen in the coding sheet in Appendices

Appendix A.

Table 9.
Predefined elements involved in completing sandwich goal

	Point awarded
2 sides of bread used	
At least 1 side was buttered	
(Vegetarian) ham included	
Cheese included	
Tomato included	
Ketchup included	
Sandwich cut in half/quarter	
Placed on plate	
Napkin on or next to plate	
Cutlery laid out	

Quality of completed goals. For any goal that was scored as either completed or partially completed, the quality with which the participant executed the goal was rated as 1 (poor), 2 (fair) or 3 (good) with respect to both the quality of the final prepared items and the strategy that was employed in the service of the goal. For example, for the sandwich goal, one point was given if a sandwich was hastily assembled or presented poorly (i.e., had ingredients missing or ingredients not evenly distributed), two points were given if the sandwich was made to a reasonable standard (i.e., contained correct ingredients but may be a little sloppy in appearance or rushed in preparation) or three points if the sandwich was made to an excellent standard. Details of the specific criteria for each goal are available in Appendices

Appendix A.

Executive error score. A checklist was created of possible errors that could be made for each of the following task goals: preparing cups of tea, preparing sandwiches, preparing folders, setting the table, the overall task and omitting any set of goals entirely. These checklists contained a list of behaviours that could be displayed that would indicate that a failure of EF had occurred. Participants received one point for each error that was present in their administration of the task and a proportion of possible errors was calculated for each set of goals. If a participant did not complete an entire task goal (e.g., failing to prepare any sandwiches), no executive errors were recorded for this goal specifically. For the goals where executive errors were recorded, a mean of proportions of possible errors was then calculated.

The possible errors were either (i) errors of commission, including any actions completed that demonstrated a failure of EF, for instance, preparing more items than needed, creating too many sandwiches or making a sandwich with chocolate spread; and (ii) errors of omission, including any behaviours that were not completed that demonstrated a failure of EF, for instance, the participant not including themselves, neglecting a set of goals entirely or preparing too few sandwiches (see Appendices

Appendix A for the complete list). Participants' behaviour was rated here even if s/he later corrected the error. For example, if the participant made more sandwiches than were necessary, and thus indicated failure of executive function, they would receive 1 point, even if the participant later realised and discarded them.

Results

Data screening

To begin, I reverse coded the proportion of dysexecutive errors committed, such that higher scores on every variable indicated better performance (the reverse coded score was used in all subsequent analysis). I assessed the distribution of each dependent variable for normality using the Shapiro-Wilk Test. The proportion of goals completed, and the involvedness score were normally distributed. The quality of execution and dysexecutive errors differed significantly from the normal distribution and Tukey's Ladder of Powers was applied using the *rcompanion* package in R (Mangiafico, 2016). The dysexecutive errors score was successfully transformed to approximate the normal distribution. Transformations applied to the quality of completed goals score did not improve its normality and so non-parametric statistics were used to compare groups on this score. To assess inter-rater reliability, videos for five participants (~6.5% of the sample) were randomly selected and second coded by an independent researcher, blind to diagnostic group and trained on the coding scheme. A significant, positive intra-class correlation coefficient was found between each score for the two rates, ICC = 0.90.

All four variables were highly inter-correlated in both the autistic and non-autistic samples (see

Figure *14*). I therefore created a composite score to characterise overall performance on the MMEA by calculating the mean of the proportion of goals completed, the quality of goals executed, the involvedness score for goals completed and the reverse coded and transformed-to-normal proportion of possible executive errors committed (hereafter referred to as the MMEA composite).

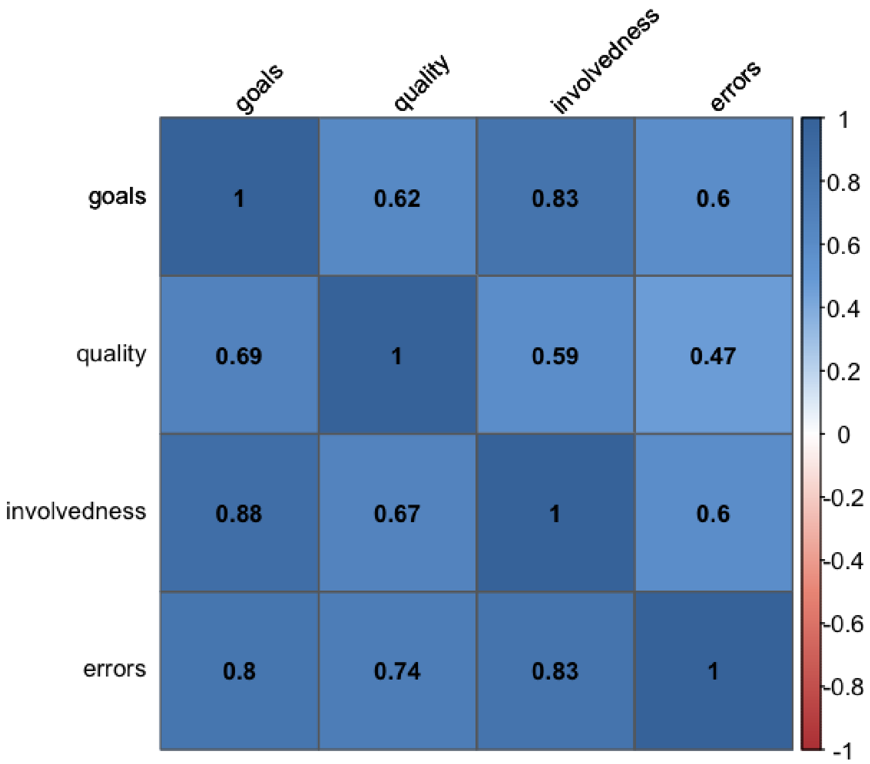


Figure 14. Pearson product-moment correlation coefficients for the relationships between the variables generated by the MMEA. The autistic sample are below the horizontal, non-autistic sample are above the horizontal. Colour and hue indicate direction and strength of the relationship; coloured boxes indicate a significant correlation. Errors = the reverse coded and transformed-to-normal proportion of possible executive errors committed. *p*-values have been adjusted for multiple comparisons.

Analysis plan

To address my first aim, to test the hypotheses that autistic participants would have more difficulties than non-autistic participants on each of the five measures generated by the MMEA (goals, involvedness, quality, executive errors and MMEA composite), I conducted a series of one-tailed independent sample *t*-tests, or Mann-Whitney *U* tests for non-normally distributed variables. I had a priori hypotheses that autistic participants would struggle on each dependent measure from the MMEA and so an unadjusted alpha is justified in these comparisons. Despite these hypotheses, a group difference on any single measure could be claimed to be evidence for less competency on the task and in-turn could in theory be classified as evidence for the existence of EF difficulties, I therefore also reported Bonferroni-Holm adjusted *p*-values which control the Type I error rate in light of the five comparisons being made (the MMEA composite, goals completed, quality, involvedness and errors).

To address my second aim, I conducted a series of Pearson product-moment correlations to determine whether individual differences in age, verbal ability, nonverbal ability, or autism symptomatology were related to task performance, indexed by the MMEA composite.

Third, I tested whether autistic participants differed from non-autistic participants with respect to the metacognitive or dispositional questions answered before and after the task by performing a series of 2 x 2 factorial ANOVAs to test for main effects of time (pre-task, post-task) and group (autistic, non-autistic) on responses. *F* tests within ANOVA have been shown to have sufficient control over Type I error rates when used with Likert scale items when the following criteria are met: (i) at least five response options are given, (ii) the underlying concept being measured is continuous and (iii) the intervals between points are approximately equal (Glass, Peckham, & Sanders, 1972).

Finally, to test whether autistic participants differed from non-autistic participants with respect to their self-appraisal upon completing the task, I conducted Spearman's *Rho* correlations between post-task appraisal scores and performance on the MMEA in the autistic and non-autistic samples separately and conducted Fisher *r*-to-*z* transformations to test whether the correlation coefficients differed significantly by group. Analyses were conducted in R (version 3.4.3; R Core Team, 2017) and RStudio (version 1.1.4; RStudio Team, 2015).

Group differences

Task performance. Table 10 shows the descriptive statistics and group comparisons for the MMEA. There were significant group differences in the proportion of goals achieved, $t(73) = 1.68$, $p =.048$, $d = 0.39$, the quality of the completed goals, $W = 989.50$, $p = .001$, $d = 0.75$, the involvedness of the completed tasks, $t(73) = 1.82$, $p =.036$, $d = 0.42$, and in MEA composite, $t(73) = 2.32$, $p = .012$, $d = 0.54$, but not in the number of executive errors committed, $t(73) = 1.70$, $p =.047$, $d = 0.32$. When the Bonferroni-Holm correction for making five comparisons is applied only the quality of the completed goals

(p = .006) and the MMEA composite (p = .046) remain significantly different across groups[5].

Table 10.
Descriptive statistics and group differences for the MMEA

	Non-autistic (n = 37)	**Autistic (n = 38)**		
	M (SD) Range	*M* (SD) Range	*p*-value unadjusted	*p*-value adjusted
Goals achieved	0.69 (0.16) 0.31 – 1.00	0.61 (0.24) 0.03 – 1.00	**.048***	.108
Involvedness score	0.61 (0.16) 0.22 - 0.95	0.53 (0.20) 0.03 - 0.85	**.036***	.108
Quality of achieved goals	0.74 (0.19) 0.33 – 1.00	0.59 (0.23) 0.27 – 1.00	**.001****	**.006****
Dysexecutive errors	0.78 (0.09) 0.61 - 0.92	0.74 (0.14) 0.36 - 0.98	.088	.108
MMEA composite score	0.65 (0.14) 0.32 - 0.87	0.56 (0.20) 0.12 - 0.89	**.012***	**.046***

Note. Means and standard deviations for the dysexecutive error score are the reverse coded but untransformed values, the *p*-values reported are based on one tailed independent samples *t*-tests conducted on reverse coded and transformed scores. Adjusted *p*-values have been adjusted for making 5 comparisons with the Bonferroni-Holm method.

Disposition toward the task. To test whether autistic and non-autistic participants' disposition toward completing the assessment differed, I conducted a series of 2 x 2 ANOVAs with time (pre-task, post-task) as the within-participants factor and group (autistic, not autistic) as the between-participant factor on participants' subjective ratings regarding task performance. First, with respect to self-reported motivation, there was a main effect of time, $F(1, 144) = 9.32$, $MSE = 0.89$, $p = .003$, $\eta^2_p = 0.61$, such that participants rated their motivation to complete the task lower before beginning the task than they did after completing the task. There was also a main effect of group, $F(1,144) = 5.55$, $MSE = 0.89$, $p = .020$, $\eta^2_p = .037$, such that autistic participants rated themselves as less motivated both pre- and post-task. There was no interaction effect, $F(1,144) = 0.37$, $MSE = 0.89$, $p = .542$, $\eta^2_p = .003$, indicating that autistic participants were not changing their reported level of motivation to a greater or lesser extent than non-autistic participants.

[5] For quality of completed goals and the MMEA composite, two additional ANOVAs including gender as a between-subjects factor revealed no significant main effects or interactions with gender.

Second, with respect to participants' appraisal of how much they would/did enjoy completing the task, there was a main effect of time, $F(1,143) = 5.84$, $MSE = 0.95$, $p = .017$, $\eta^2_p = .039$, such that participants' pre-task predictions about how much they would enjoy the task were lower than their post-task ratings of how much they enjoyed completing the task. There was also a main effect of group, $F(1,143) = 5.28$, $MSE = 0.95$, $p = .023$, $\eta^2_p = .036$, such that autistic participants said they would enjoy or did enjoy completing the task less than non-autistic participants. There was no interaction effect, $F(1,143) = 0.33$, $MSE = 0.95$, $p = .566$, $\eta^2_p = .002$.

Meta-cognition. There was neither a main effect of time, $F(1,144) = 0.31$, $MSE = 1.05$, $p = .576$, $\eta^2_p = .002$, nor of group, $F(1,144) = 3.40$, $MSE = 1.05$, $p = .067$, $\eta^2_p = .023$, on participants' ratings of how much effort the task would/did take to complete. There was also no interaction between time and group, $F(1,144) = 0.52$, $MSE = 1.05$, $p = .472$, $\eta^2_p = .004$. Furthermore, Spearman's rank-order correlations were conducted to test the relationship between participants' self-appraisal in the post-task question about how well they believed they performed, rated from 1 (not well at all) to 5 (very well) and the MMEA composite, which was significant for autistic, $r_s = .46$, $p = .004$, and non-autistic participants, $r_s = .47$, $p = .004$ (see Figure 15). That is to say, participants who performed better on the task were more likely to rate themselves as having performed well, in both groups. Furthermore, Fisher's r-to-z transformation was applied to the correlation coefficients and showed that there was no group difference in the relationship between task performance and task appraisal, $z = 0.05$, $p = .480$, both groups were equally as reliable at appraising task performance.

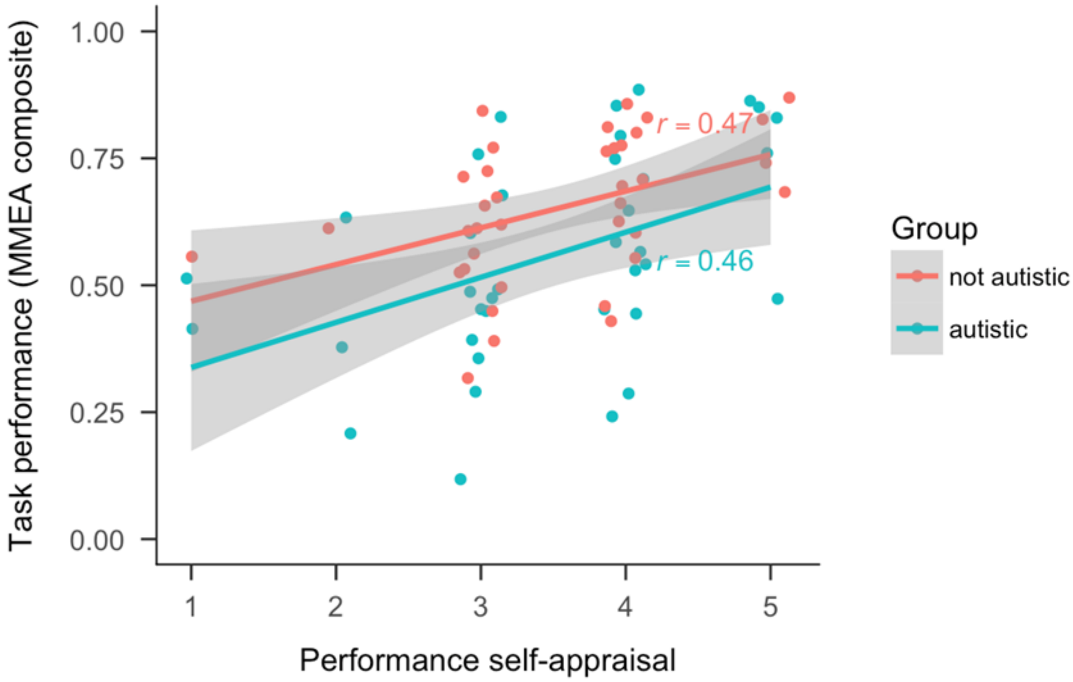


Figure 15. The relationship between participants' execution of the MMEA, as indexed by the MMEA composite, and their post-task self-appraisal of their performance from 1 (not at all well) to 5 (very well). Data jittered to avoid over plotting. Lines indicate Spearman's Rho correlation, by group.

Individual differences

The final aim of the present study was to test whether there were relationships between performance on the MMEA and on other, non-executive variables. Figure 16 shows the relationships between, the MMEA composite and: (A) verbal IQ, (B) performance IQ, (C) age and (D) autistic traits, as indexed by the SRS-2. I conducted a series of Pearson product-moment correlations for each of the above. Age was associated with the MMEA composite among autistic, $r(36) = .39$, $p = .014$, and non-autistic groups, $r(35) = .48$, $p = .003$. An r-to-z transformation showed that these relationships were not significantly different, $z = 1.55$, $p = .060$. There was a significant association between verbal IQ and performance for autistic participants only, $r(36) = .34$, $p = .038$. None of the other tested correlations were significant (all ps > .484).

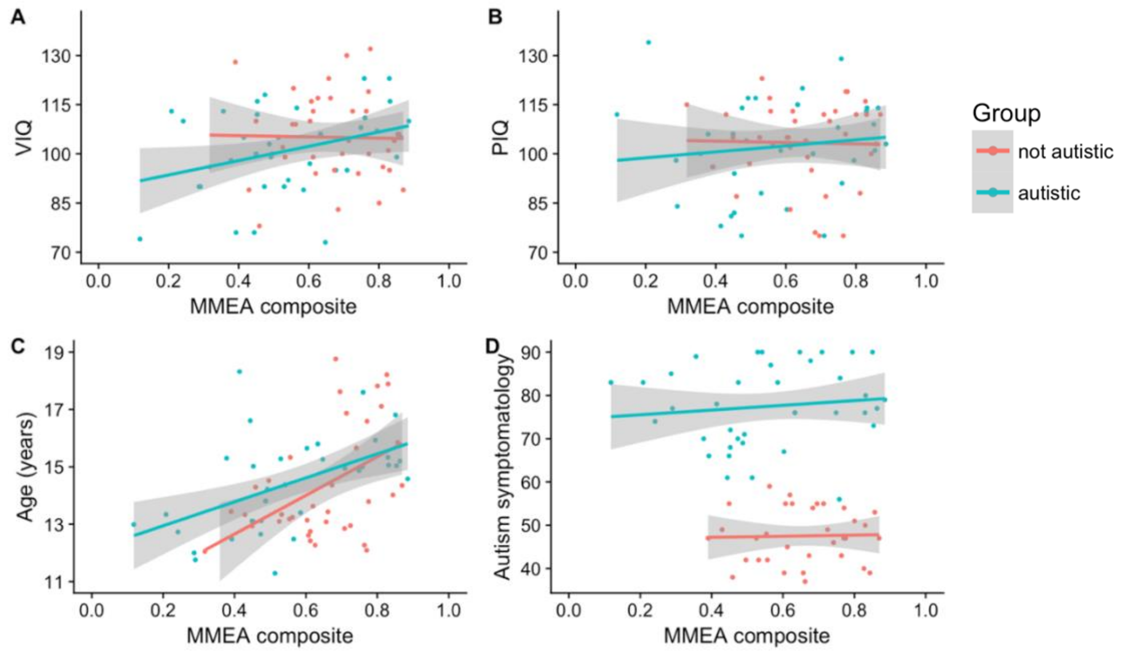


Figure 16. Relationship between performance on the composite score from the Modified Multitasking Evaluation for Adolescents (MMEA) and (A) VIQ = verbal comprehension index on the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011), (B) PIQ = perceptual reasoning index on the WASI-II, (C) age (in years) and (D) autism symptomatology, indexed by t-scores from the Social Responsiveness Scale, second edition (Constantino & Gruber, 2012).

Discussion

This study sought to modify an existing, naturalistic EF task selected because it represents the conditions over which we aim to generalise results of EF assessments, that is, executively demanding everyday tasks. This was the first time this assessment was used to compare the performance of autistic and non-autistic adolescents. As well as showing this task to be feasible with this population (although two autistic participants chose not to participate), I found group differences in the proportion of goals completed, the quality of the completed goals, the involvedness of completed goals and on the composite performance score, such that autistic participants, on average, under-performed relative to non-autistic participants. Once corrections had been made for multiple comparisons, group differences on the quality of executed goals and on the overall composite remained significant.

The task used was representative of a situation where participants, within the age range assessed, might be expected to deploy executive abilities with respect to both its form (i.e., dealing with strong sensory input, moving in veridical space, engaging in open-ended problem solving) and context (completing errands within a time limit, engaging in a self-guided and concrete exercise). These aspects of the task serve to increase our confidence that the EF difficulties often reported in the literature do extend beyond classic laboratory assessment paradigms, which represent neither the form nor context of real-world activities particularly well. This result is consistent with the findings of Rajendran et al. (2011) and Altgassen et al. (2012), who showed autistic adolescents and adults struggle in virtual and naturalistic EF tasks, respectively. The current study extends these findings by providing evidence in a veridical rather than a virtual task and by finding differences in overall task competency rather than simply in prospective memory-related errors.

Autistic participants' performance was rated as being of poorer quality than the non-autistic participants, suggesting that the goals that were attempted, either involved less efficient strategies (e.g., walking back and forth, multiple times, between the preparation area and the table being set) or that the finished products were of poorer quality (e.g., the sandwiches were poorly assembled). This effect size estimate was larger than any of the effect size estimates between groups on the NIH-EXAMINER battery (described in Chapter 2) were. One possible explanation, as suggested by White et al. (2009) and consistent with the difficulties participants showed on open-ended tasks in Chapter 2, is that in less constrained environments, autistic people have more difficulties than in the artificially-structured standard EF assessments. The group difference on this metric of performance supports my hypothesis that classic EF assessment paradigms underestimate the difficulties autistic people have in everyday life. It must be noted; however, that quality of performance was the most subjective metric of performance, which may have driven, at least in part, the large effect size, especially considering that the videos were not coded by a researcher blinded to group membership.

Interestingly, autistic participants did not commit a higher proportion of the pre-specified dysexecutive errors that were possible during task execution. This finding speaks against my prediction that autistic people would make more dysexecutive errors, on account of demands placed on planning and cognitive flexibility abilities, which autistic people are well documented to struggle with within laboratory assessments (Geurts, Corbett, et al., 2009; Olde Dubbelink & Geurts, 2017). It remains probable that the lower scores achieved by autistic participants on overall task performance and on the quality of goal execution were related to more subtle executive difficulties that were not included in the list of dysexecutive errors, which included the most profound possible failures of EF. The task, by design, involved the assessment of EF in the context of many non-executive factors (e.g., social, sensory, and affective factors) and it is therefore not clear the extent to which executive and non-executive factors were implicated in the lower performance scores.

Motor coordination is one such non-executive factor that had a bearing on task performance that warrants further investigation. For instance, fine motor co-ordination skills had an impact on participants' ability to complete some of the highly dexterous aspects of this task, such as stapling sheets together and buttering bread. Steps were taken to minimise the requirements for these skills in this task, for example by ensuring that all of the packets were open before participants began completing the task. Additionally, gross motor skills were also indirectly assessed in this assessment as participants needed to move around the room in order to set the table, to use different work surfaces and to return to previous parts of the room to complete time-sensitive subtasks, such as returning to a boiled kettle. It remains possible that autistic and non-autistic participants differed with respect to fine-motor and/or gross motor abilities, as has been shown previously (Fournier, Hass, Naik, Lodha, & Cauraugh, 2010), and this may have explained some of the group differences reported herein.

Finally, although the task was selected because a participant could complete it alone, thereby reducing any social demands placed on participants, some sociocognitive demands likely remained within the task context. For instance, the task was video recorded by an experimenter who was physically present in the room. The presence of an experimenter may have induced social anxiety for some participants, and it is plausible that autistic participants were more likely to experience this anxiety (Bellini, 2006). State anxiety during the assessment was not assessed and so the group differences found may well be compounded by differences in induced anxiety levels while completing the task. Similarly, the participants were asked to prepare for an imagined social event. This effect of this given context may have differed across groups. For example, the samples may have differed with respect to their social motivation to fulfilling the examiner's demand in the knowledge that the materials were not needed for a real event. Further, the samples may have differed in the extent to which they would have been motivated by this type of social situation if it were it to have been a real study group. These social aspects surrounding this task may also have confounded the group differences in task performance. While the socio-affective elements of the task render it more difficult to make causal claims about group differences, they do more closely resemble the conditions under which tasks are completed outside of the laboratory.

More research is required that employs an experimental design to control for each of these non-executive factors to determine with greater accuracy if differences on a task such as this one was related to group differences in EF over and above these non-executive factors.

It was likely that the amount of experience a participant had completing activities similar to those assessed on this task affected task performance. Unfortunately, the amount of experience participants had on these activities was unmeasured in the present study. The role previous experience plays in task performance has implications both for the measurement of EF ability and for our theoretical understanding of EF development. First, EF can only be truly measured by a task that is novel and when a participant has not automatized a response to task demands (Burgess et al., 2006). The present task was not consistently novel to all participants. Consequently, the extent to which it was executively taxing differed for different participants. Future investigations should ask participants to report their experience with similar tasks, as was done by Altgassen et al. (2012).

Second, with respect to our theoretical understanding of EF development, it is unclear if participants' experience with similar activities differed between groups or interacted with task competence. The neuroconstructivism account of development suggests that if autistic people struggle markedly with EF early in development then parents and teachers may change the way they are treated, reduce the executive demands in their home and educational environments which, critically, limits the opportunities they have to exercise and develop these abilities (Karmiloff-Smith, 2009). According to Karmiloff-Smith (2009), treating those with difficulties differently, serves to compound difficulties through development. In the current study, increasing age was similarly associated with higher task performance scores in both groups, suggesting this not to be the case. Future studies should assess participants who span more diverse stages of development, by for instance, including young and older adults to test whether executive difficulties abate or whether they are compounded through development, as was suggested by Karmiloff-Smith (1998).

Autistic participants rated themselves as being less motivated and as enjoying this task less than non-autistic participants did. Yet, autistic participants committed a comparable proportion of executive errors to non-autistic participants did. It is possible that the group differences on overall task performance were related to the different disposition autistic participants had toward completing the task, rather than to differences in EF ability. The group differences in disposition toward this task has important implications for the ethological validity of naturalistic EF assessment, in particular, and for interpreting the results of cognitive assessments, in general. It is unclear whether different ratings reflect differing dispositions toward completing the task because of demand characteristics, like social desirability or inferring implicit information (S. J. White, 2013) or if differences are intertwined with participants previous experience of similar tasks and their awareness of the difficulties that such tasks present for them. It is also unclear whether these differences are specific to a naturalistic task or that they may also be implicated in results garnered from standard EF tasks.

There were no group differences in participants' ability to appraise their overall performance, which failed to support my hypothesis that autistic youth would lack insight into the difficulties they had completing this activity. This diverges from previous accounts that held autistic young people to be unreliable informants about their own cognitive abilities (U. Frith & Happé, 1999), which may have resulted in the widespread application of informant-reported but not self-reported EF questionnaires. The finding that cognitively able autistic young do not lack insight also demonstrates a critically important difference between autistic people and other populations known to have EF difficulties (Burgess et al., 1998). An implication for the assessment of EF ability follows: we must do more to include the perspectives of young autistic people when assessing their EF ability and not rely solely on proxy informants in interview and questionnaire studies.

In sum, the current study demonstrated that cognitively able autistic adolescents show EF difficulties, on average, relative to their non-autistic peers in a representatively-designed EF task, thus strengthening our confidence that the difficulties measured in Chapter 2, and in the literature more broadly, manifest in the lives of young autistic people. Representativeness is one criterion by which the ecological validity of an assessment can be appraised but it does not provide sufficient evidence to conclude ecologically validity. Chapter 4 will therefore directly test the generalisability of performance on both the NIH-EXAMINER and the MMEA to other measures of everyday functioning.

Chapter 4

The generalisability of executive function assessments to everyday functioning in autistic and non-autistic adolescents

Introduction

Chapter 2 demonstrated that a sample of autistic adolescents underperformed relative to their non-autistic peers on a composite measure derived from a wide-ranging battery of neuropsychological executive function (EF) assessments. It also showed, however, that the component tasks within that battery yielded, for the most part, no significant group differences or effect sizes that were smaller than anticipated. Chapter 3 demonstrated that a sample of autistic adolescents, who largely overlapped with the sample from Chapter 2, underperformed relative to their non-autistic peers on an executively-demanding naturalistic assessment that was also more representative of everyday life. Despite overall group differences on this measure, autistic adolescents made no more executive errors while completing the task than their non-autistic counterparts. It remains unknown whether the magnitude or nature of the difficulties autistic people showed on both EF assessments are linked to the difficulties they encounter in daily life.

This Chapter describes my third empirical study, which aimed to address this issue directly, by testing whether performance on a neuropsychology battery of EF assessments (NIH-EXAMINER) and competence on an executively-demanding everyday task (the MMEA) predict participants' self-reported (dis)abilities in several domains of everyday life. I begin by introducing the importance of assessing the generalisability of EF assessment before outlining the literature in which the generalisability of EF assessment has been specifically tested with autistic people.

Generalisability of EF assessment

Showing that autistic people have difficulties on representatively-designed tasks, such as the one described in Chapter 3, as well as on systematically-designed tasks, such as those described in Chapter 2, strengthens our ability to infer that the difficulties apparent in these assessments manifest in the daily lives of some autistic people. Demonstrating that autistic people have difficulties on a representatively-designed task provides evidence for the ecological validity of EF assessment according to the verisimilitude criterion, but as discussed in Chapter 1 this may not be sufficient evidence that these difficulties necessarily affect the everyday lives of autistic people (e.g., see Vaskinn, Sergi, & Green, 2009). Another approach to determining the ecological validity of an assessment, the veridicality criterion, involves testing whether performance on the assessment of interest is related to other real-world measures, that is, that the task has generalisability (Burgess et al., 2006; Kvavilashvili & Ellis, 2004) or as it is referred to by others, *veridicality* (Chaytor & Schmitter-Edgecombe, 2003; Franzen & Wilhelm, 1996).

The generalisability of an assessment is necessarily influenced by the real-world measure against which it is assessed. When assessing the veridicality of EF in other populations, some investigations have used general measures, such as academic achievement (Blair & Razza, 2007), employment status (Bayless, Varney, & Roberts, 1989) and occupational performance (Barkley & Fischer, 2011; Kibby, Schmitter-Edgecombe, & Long, 1998). Other investigations compared performance on EF assessments with questionnaires tapping participants' adaptive behaviour (C. Clark, Prior, & Kinsella, 2002; Gilotty et al., 2002) or everyday EF ability, measured by, for example, the Dysexecutive Questionnaire (Burgess et al., 1998; Wood & Liossi, 2006), the Behaviour Rating Inventory of Executive Function (Vriezen & Pigott, 2016) and the Deficits in EF scale (Barkley & Fischer, 2011).

The abovementioned studies drew mixed conclusions about the generalisability of EF assessment. On the one hand, participants' ability to construct an object using wooden wheels and plastic sticks from a child's toy set, and the complexity of the constructed object, distinguished between patients who returned or failed to return to work following head injury (Bayless et al., 1989). Similarly, performance on the Brixton test did correlate with questionnaire reports of executive difficulties (Wood & Liossi, 2006).

On the other hand, there are numerous studies showing limited generalisability of EF. For example, for adults with attention deficit hyperactivity disorder (ADHD), self- and informant-ratings of EF ability were largely unrelated to performance on EF assessments but occupational functioning was more closely linked to the questionnaire ratings than to performance on the EF tests (Barkley & Fischer, 2011). Another study, this time assessing generalisability of assessment against a rating scale about everyday EF ability, found that performance on the Wisconsin Card Sorting Test, a verbal fluency task and the Trail Making Test and were not related to any scale from the parent-reported BRIEF in children with traumatic brain injuries (Vriezen & Pigott, 2016). Similarly, in another sample with acquired head injuries, there were no relationships between performance on the Hayling Test, the Zoo Map and Key Search tasks, taken from the Behavioural Assessment of the Dysexecutive Syndrome (BADS), and informant-reported Dysexecutive Questionnaire scores (Wood & Liossi, 2006). The relationship between laboratory assessments and everyday measures can be population-specific, for example, overall performance on the BADS was related to informant-reported EF difficulties in patients with head injuries, but not those with schizophrenia (Evans, Chua, Mckenna, & Wilson, 1997).

With respect to the NIH-EXAMINER battery, one study with a sample of mixed neurological patients demonstrated that the global executive composite score was related to both informants' perception of their real-world EF competencies, measured by the Frontal Systems Behaviour Scale (FrsBe; Grace & Malloy, 2001), and to variation in brain volume in regions critical for executive control (Possin, LaMarre, Wood, Mungas, & Kramer, 2014).

The composite score remained a significant predictor of informant reports even after two more widely used EF measures (the Trail-making and Stroop tasks) were included as covariates. Similarly, individual differences on the global executive composite was related to ratings of everyday EF ability, measured by the FrsBe for adults and by the Behaviour Rating Inventory of Executive Function (BRIEF; Gioia, Guy, Isquith, & Kenworthy, 1996) for children in the initial validation sample (Kramer et al., 2013). Thus, demonstrating that, at least with respect to samples of neuropsychological patients, results from the NIH-EXAMINER battery can be generalised to other measures of real-world EF and so meets the veridicality criterion for demonstrating ecological validity. With respect to the MMEA, the previous version of this assessment was related a measure of informant reported real-world functioning in a sample of adolescents with to 22q11.2 Deletion Syndrome (M. Schneider et al., 2016). Task veridicality can, however, be population specific (Kenworthy et al., 2008) and so veridicality must also be demonstrated with an autistic sample.

In reviewing the ecological validity of cognitive assessments, Chaytor and Schmitter-Edgecombe (2003) found that the ecological validity of EF tasks was poorer than that of other cognitive constructs reviewed, such as memory. In this review, the authors concluded that assessments with greater verisimilitude (i.e., representatively-designed assessments) had greater veridicality (i.e., were more likely to generalise to and be statistically related to other, real-world measures) but this conclusion, according to the authors, was hampered by the small number of studies that assessed the generalisability of any cognitive assessment. Furthermore, the extent to which performance on tasks such as the BADS relate to other, real-world measures appears to be population specific (Evans et al., 1997; Kenworthy et al., 2008). It is therefore necessary to determine whether performance on EF tasks relates to other real-world measures for autistic people and, if they do, to test whether the generalisability of the assessments is comparable for autistic and non-autistic people.

Veridicality of EF assessment in autism

Remarkably few studies have assessed the generalisability either of standard EF assessments or of representatively-designed EF tasks specifically created for autistic samples. With respect to neuropsychological or experimental measures of EF, some studies have failed to demonstrate a link between performance on EF tasks and everyday skills (e.g., Davids, Groen, Berg, Tucha, & van Balkom, 2016; Strang et al., 2017). Other studies, however, have shown results that are more positive. In one cross-sectional study, Pellicano et al. (2017) demonstrated that the reverse reward task (measuring inhibition) and the Corsi blocks task (measuring spatial working memory ability) predicted unique variance in a standardised assessment of school readiness in autistic and non-autistic pre-school children over and above the variability explained by age, ability and diagnostic status (Pellicano et al., 2017). Other studies have shown that individual differences in children's performance on experimental EF measures correlate with a parent-report measure of adaptive functioning both cross-sectionally (Happé, Booth, et al., 2006; Panerai et al., 2014) and longitudinally (Berger, Aerts, Spaendonck, Cools, & Teunisse, 2003), even up to 12 years later (Kenny et al., 2018).

With respect to representatively-designed EF tasks, one study investigated the relationship between performance on the Battersea Multitask paradigm and parent-reported scores on the BRIEF for autistic boys and found that once corrections for multiple comparisons were applied no relationships remained significant (Mackinlay et al., 2006). In another study, autistic adults completed the Dresden breakfast task and rated their own EF ability on the Dysexecutive Questionnaire (Altgassen et al., 2012). Participants' time-based, but not event-based, prospective errors were correlated with their questionnaire responses, but the authors did not test whether an overall metric of task execution correlated with participants' self-reported EF ability. An unpublished measure of EF, called the Ecologically valid Test of Executive Dysfunction, which was designed to address some concerns of ecological validity and to be appropriate for the clinical assessment of autistic children and young people, has been compared with parent-reported BRIEF scores (Bristow, 2016; Pullinger, 2017). There were relationships between parental accounts of executive abilities among their autistic children (8 - 12 years) and participants' performance on a drawing task tapping planning ability and a story re-telling task tapping working memory ability (Pullinger, 2017). There were, however, no relationships between parent ratings and performance on any outcome variables from the remaining four activities within the battery (Bristow, 2016).

The current study

The aims were threefold. First, I aimed to assess the convergent validity the MMEA by assessing the degree to which performance on the NIH-EXAMINER and on the MMEA were related, separately for the autistic and non-autistic samples. If both of these assessments are indeed measures of the same underlying construct, i.e. EF, then performance on both assessments should be related. I predicted therefore that performance on the two EF assessments would be related in each group, demonstrating evidence convergent validity for the MMEA. I also predicted that the magnitude of the relationship would be smaller for the autistic sample given the non-executive factors, such as the sensory, social and motor demands of the task would likely have a greater effect on performance among autistic participants. I also assessed whether there was an interaction between group (autistic, not autistic) and assessment type (NIH-EXAMINER, MMEA). I predicted that the difference between performance on the MMEA and the NIH-EXMAINER would be larger for autistic participants than for non-autistic participants based on autistic participants' difficulty with less-structured tasks (S. J. White et al., 2009) and the relatively unstructured nature of the MMEA (M. Schneider et al., 2016).

Second, I sought to test the generalisability of both a systematically-designed assessment (the NIH-EXAMINER battery) and a representatively-designed assessment (the Modified Multitasking Evaluation for Adolescents [MMEA]) of EF. Specifically, I sought to test whether EF difficulties were related to variability in domains of life with which autistic people often report experiencing difficulties, including everyday EF difficulties, quality of life, trait anxiety and self-perceptions of disability. While previous research has assessed the relationship between parental reports of EF and quality of life (e.g., de Vries & Geurts, 2015), mental health difficulties (e.g., R. A. Lawson et al., 2015) and functional skills (e.g., Panerai et al., 2014), performance-based measures of EF have not been validated against such a diverse range of metrics of everyday functional adjustment.

I was interested in understanding whether individual differences on the performance-based measures were related to participants' own appraisal of their everyday functional adjustment and so self-reported questionnaires were used. To address this aim, I first assessed whether the current sample showed difficulties relative to the non-autistic participants on self-reported measures of each of these domains. Next, I tested whether individual differences on scores on either measure of EF (the NIH-EXAMINER battery and the MMEA) were related to variation in scores on each domain of everyday functional adjustment assessed, for the autistic and non-autistic samples, separately. Based on autistic participants' difficulties with less-structured tasks, I predicted that performance on the MMEA would have greater predictive utility over other real-world measures than would performance on the NIH-EXAMINER.

Method

Participants

Sixty-four adolescents aged between 11 and 19 years (M = 14.53, SD = 2.04) took part in this study, 33 autistic (11 female, 22 male) and 31 non-autistic (12 female, 19 male) participants completed all measures. There were eight additional participants who participated only in the MMEA and did not complete the NIH-EXAMINER, four additional participants who completed only the NIH-EXAMINER and did not complete the MMEA and three further participants for whom self-reported questionnaires could not be obtained, despite repeated attempts, data for these participants are not presented here.

Inclusion and exclusion criteria were the same as those reported in Chapters 2 and 3. Method of recruitment, participants' co-occurring conditions, medication use, and participant ethnicities are listed in Table 11 below. One autistic participant and six of the non-autistic participants did not speak English as their first language. Parents reported on participants' ethnicity. A Fisher's exact test revealed that these differences were not significant, p = .097 (see Table 11).

Table 11.
Demographic information for participants who completed the NIH-EXAMINER battery, the MMEA and the self-reported questionnaires, by group

	Non-autistic (*n* = 31)	**Autistic (*n* = 33)**
	n	*n*
Recruited through		
Public engagement and research participation event	30	24
Community contacts	1	4
School-research partnership	0	5
Co-occurring conditions		
Attention deficit hyperactivity disorder	0	9
Dyslexia	0	9
Developmental coordination disorder	0	7
Sensory processing disorder	0	3
Irlen Syndrome	0	1
Obsessive compulsive disorder	0	1
Complex language disorder	0	0
Medication		
ADHD	0	6
Antipsychotics	0	1
Sleep inducing	0	4
Antiepileptic	0	0
Antidepressants	1	2
Non-psychoactive	2	1
Ethnicity		
Any White background	14	23
Any Asian background	8	1
Any Black background	3	3
Any mixed background	3	4
Other ethnic group	1	1
Missing or prefer not to say	2	1

Note. Any White background = White British, White Irish or any other White background; Any Black background = Black British, Black African, Black Caribbean or any other Black background; any Asian background = Chinese, Indian, Pakistani, Bangladeshi or any other Asian background; Mixed/multiple ethnic groups = Mixed White and Asian, Mixed White and Black African, Mixed White & Black Caribbean, Any other Mixed background.

Developmental and background data for the final sample can be seen in Table 12. Data were excluded for one additional autistic participant who obtained an SRS-2 *t*-score of 56 which fell below the threshold indicative of being autistic (a *t*-score equal to or above 60; Constantino & Gruber 2012; ADOS-2 data were unavailable for this participant). One autistic participant obtained an SRS-2 *t*-score of 59, but nevertheless met criteria for an autism spectrum condition on the Autism Diagnostic Observation Schedule, second edition (ADOS-2; Lord, Rutter, et al., 2012) and so their data were retained for further analyses. Data for two additional non-autistic participants (both female) were excluded because they obtained SRS-2 *t*-scores of 65 and 69, which fell above the threshold indicative of being on the autism spectrum. All participants were cognitively able as indicated by having verbal, nonverbal and full-scale IQ scores of 70 or above (see Table 12). Two additional participants' data were excluded because at least one of these scores were missing. An additional five autistic participants were excluded for obtaining IQ scores below this threshold. The final samples were group-matched by age, $t(62) = 0.90$, $p = .369$, $d = 0.23$, verbal IQ , $t(62) = 1.13$, $p = .263$, $d = 0.63$, nonverbal IQ, $t(62) = 0.07$, $p = .948$, $d = 0.02$ and by gender distribution, χ^2 (1, $n = 64$) = 0.04, $p = 0.85$, $\phi = 0.06$.

There were no significant differences between the autistic and non-autistic participants in level of parental/caregiver education, $t(57) = 0.72$, $p = .477$, $d = 0.19$. A decile rank of deprivation and number of years of post-16 education was calculated for a parent of each participant, as in Chapter 2. There was a significant group difference on the decile rank of deprivation, $t(54) = 2.16$, $p = .035$, $d = 0.58$. There were no differences on the SES composite, $t(49) = 0.32$, $p = .753$, $d = 0.09$, however.

Table 12.

Descriptive statistics for the developmental and background variables for participants who completed the NIH-EXAMINER battery, the MMEA and the self-reported questionnaires, by group

	Non-autistic (*n* = 31)	**Autistic (*n* = 33)**	
	M (SD) Range	*M* (SD) Range	*p*-value
Developmental variables			
Gender (M:F)	20:11	21:12	.851
Age (in years)	14.29 (2.07)	14.75 (2.02)	.369
	12 - 18	11 - 19	
Full-scale IQ	105.52 (12.34)	103.00 14.67	.462
	77 - 130	76 - 132	
Verbal IQ	105.23 (13.86)	101.48 (12.61)	.263
	78 - 132	73 - 123	
Nonverbal IQ	103.90 (13.47)	104.18 (19.88)	.948
	75 - 123	75 - 154	
SRS-2	48.00 (6.04)[a]	77.12 (9.45)	**< .001**
	38 - 59	59 – 90[b]	
ADOS-2 severity score[c]	- -	5.61 (2.64)[d]	-
	-	2 - 10[e]	
SES of primary caregiver			
Years of post-16 education	5.63 (4.99)[f]	4.88 (3.00)[g]	.477
	-1 – 19[h]	0 - 12	
IMD decile	5.19 (3.33)[i]	6.97 (2.82)[j]	**.035**
	1 - 10	1 - 10	
SES composite score	-0.03 (0.77)[k]	0.03 (0.64)[l]	.753
	-1.20 – 1.13	-1.17 – 1.33	

Note. Full-scale IQ = Full-scale IQ, 4 subtest version, derived from the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011), where mean score is 100 and standard deviation is 15; Verbal IQ = Verbal comprehension index derived from the WASI-II (Wechsler, 2011), where mean score is 100 and standard deviation is 15; Nonverbal IQ = Perceptual reasoning index derived from the WASI-II (Wechsler, 2011); SRS-2 = *t*-score, calculated separately, by gender from the Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012); ADOS-2 = the Autism Diagnostic Observation Schedule, second edition (Lord, Rutter, et al., 2012); SES = Socioeconomic status; IMD = Index of Multiple Deprivation (T. Smith et al., 2015); Composite SES score = mean of z-scores derived from the years of post-16 education and the IMD decile rank

[a] $n = 27$. [b] one participant fell below the threshold indicative of an autism diagnosis (a *t*-score of 60) but were retained in the analysis because they received a score above the threshold indicative of an autism diagnosis on the ADOS-2. [c] The ADOS-2 was only administered with participants already diagnosed with autism. [d] $n = 18$, The ADOS-2 was not administered to all autistic participants because it was not an included measure in the planned protocol for this study due to time limitations. Data are reported here in cases where ADOS-2 data were available because participants took part in other research studies and consented for their data to be used in this way. [e] Two autistic participants received ADOS-2 severity scores below the threshold indicative of an autism diagnosis (a severity score of 2 or below) but were retained in the analysis because they received a score above the threshold indicative of an autism diagnosis on the SRS-2. [f] $n = 27$. [g] $n = 32$. [h] One parent reported leaving full time education aged 15 years and so received a score of -1 when number of years of post-16 education was calculated. [i] $n = 27$. [j] $n = 29$. [k] $n = 23$. [l] $n = 28$.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Measures

The NIH-EXAMINER battery. Participants completed all subtests of the NIH-EXAMINER battery and the executive composite was calculated using Item Response Theory using the script provided by Kramer et al. (Kramer et al., 2013). The executive composite score from the battery as described in Chapter 2, according to manual instructions, was used in this study as the measure of systematically-designed EF.

The Modified Multitasking Evaluation for Adolescents (MMEA). All participants completed the MMEA, as described in Chapter 2. The MMEA composite score was used here as a measure of representatively-designed EF.

The Behaviour Rating Inventory of Executive Functioning. Participants' everyday EF ability was assessed using the self-report version of the Behaviour Rating Inventory of Executive Function, second edition (BRIEF-2; Gioia, Isquith, Guy, & Kenworthy, 2015) which assesses the behavioural manifestation of EF difficulties. The BRIEF-2 has demonstrated acceptable reliability, and convergent and discriminant validity have been established (Gioia et al., 2015). Respondents were required to rate on a 3-point Likert scale, range from 1 (Never) to 3 (Often) how frequently they had difficulties in the past month on each item with regard to initiation, emotional control, shifting, inhibition, organisation/planning, organisation of materials, working memory and self-monitoring. Responses to each item were summed to produce an overall Global Executive Composite, as per the guidelines in the manual[6]. A parent-reported version of this measure has been used extensively to measure informant ratings of EF in autistic samples (A. S. Chan et al., 2009; de Vries & Geurts, 2015; Kalbfleisch & Loughan, 2012; Kenworthy, Black, & Wallace, 2005; Lynch et al., 2017; Pugliese et al., 2015, 2016; Semrud-Clikeman et al., 2010). An adult, self-reported version has been used with autistic samples (Davids et al., 2016; Dijkhuis, Ziermans, Van Rijn, Staal, & Swaab, 2017; Wallace et al., 2016). This is the first study to use the self-reported version of this measure with autistic adolescents. Higher scores indicated more executive function problems. Total possible raw score was 165. In the current sample, the scale showed excellent internal consistency in both autistic (α = 0.95) and non-autistic (α = 0.94) young people.

[6] BRIEF-2 subdomain scores were not considered to preserve statistical power and for parity with the composite variables that were generated from the NIH-EXAMINER and MMEA that it is being compared with

The Pediatric Quality of Life Inventory. Quality of life (QoL) was assessed using the self-report core scales of The Pediatric Quality of Life Inventory, fourth edition (PedsQL 4.0; Varni, Seid, & Kurtin, 2001) and the general well-being supplement for teenagers. The PedsQL 4.0 core scales required respondents to report, on a 5-point Likert scale ranging from 0 (never) to 4 (almost always), how often within the last month they have had difficulties on questions related to physical, emotional, social, and school-related quality of life. The PedsQL 4.0 general well-being supplement required respondents to rate on a 5-point Likert scale ranging from 0 (never) to 4 (almost always) how often within the past month they felt that positively-phrased statements about general wellbeing sound like them. Scores from the generic core scales were rescaled to range between 0 to 100 according to Varni et al. (Varni et al., 2001), where 0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0. The scores from the well-being supplement were re-scaled in the same way here and reverse coded, such that higher scores were indicative of better QoL. A mean of all generic core scales and the well-being supplement was calculated[7]. Higher PedsQL scores reflected better QoL. The total possible score was 100. The self-reported PedsQL 4.0 has been previously used with autistic children (Bastiaansen, Koot, Ferdinand, & Verhulst, 2004; Potvin, Snider, Prelock, Wood-Dauphinee, & Kehayia, 2015) and adolescents (Shipman, Sheldrick, & Perrin, 2011). In the current sample, the scale showed excellent internal consistency in both autistic (α = 0.92) and non-autistic (α = 0.94) young people.

The Spence Children's Anxiety Scale. Participants' anxiety symptoms were assessed using The Spence Children's Anxiety Scale (SCAS; Spence, 1998), which has been used extensively with autistic children (Boulter, Freeston, South, & Rodgers, 2014; Neil, Olsson, & Pellicano, 2016; Wigham, Rodgers, South, McConachie, & Freeston, 2015). The SCAS is a 44 item self-report measure of children and adolescent's anxiety, which included 38 items about anxiety and six positive filler items. Respondents were required to rate how

[7] PedsQL subdomain scores were not used to preserve statistical power

frequently they experienced each item on a 4-point Likert scale ranging from 0 (never) – 3 (always). The responses from the 38 items pertaining to anxiety were summed to create a total raw SCAS score[8]. Higher scores reflected greater trait anxiety symptom severity. The total possible score was 114. In the current samples, the scale showed internal consistency that was excellent for both for autistic ($\alpha = 0.92$) and non-autistic participants ($\alpha = 0.94$).

The World Health Organization Disability Assessment Schedule. Participants reported on their perceived disabilities using the self-reported version of the World Health Organization Disability Assessment Schedule, second edition (WHODAS 2.0; Garin et al., 2010). The WHODAS 2.0 has previously been used with autistic adults (Piening et al., 2018; L. Schmidt et al., 2015). The WHODAS 2.0 was originally designed for adults, although removing one item relating to sexual activity was removed revealed that the 35-item version retained satisfactory internal reliability and factorial validity when used with adolescents (Hu, Zang, & Li, 2012). I therefore followed the same procedure here. In the WHODAS 2.0, respondents were required to rate how much difficulty they have had with each item in the last 30 days on a 5-point Likert scale ranging from 0 (none) to 4 (extreme or cannot do). The items relate to cognition, mobility, self-care, getting along with people, life activities and participation in society. There were a different number of items contributing to each domain and to account for these differences, an item response theory algorithm provided by the WHO, was used to generate an overall total score, where higher scores reflected greater disability[9]. The WHODAS 2.0 has demonstrated internal reliability (Cronbach's alpha has ranged from α=.79 to α=.98), it has a stable factor structure, shows high test-retest reliability and good concurrent validity (Garin et al., 2010). The total possible score was 100. In the current samples, the scale showed internal consistency that was acceptable ($\alpha = 0.73$) for autistic participants and good ($\alpha = 0.84$) for non-autistic participants.

[8] SCAS subdomain scores were not used to preserve statistical power
[9] WHODAS 2.0 subdomain scores were not used to preserve statistical power

Results

Data screening

Regarding the EF measures, I began by calculating the MMEA composite, as described in Chapter 2. To this end, I reverse coded the executive errors score, such that higher scores on every variable generated from the MMEA indicated better performance. Tukey's Ladder of Powers were then applied to successfully transform the involvedness and executive errors scores (the quality score remained significantly different to the normal distribution and so raw scores were used for this variable for subsequent analysis) using the *rcompanion* package in R (Mangiafico, 2016). Next, I created an overall MMEA composite score by calculating the mean of the proportion of goals completed, the quality of executed goals, the transformed-to-normal involvedness score and the reverse coded and transformed-to-normal executive error score.

Regarding the measures of everyday adjustment, only the PedsQL 4.0 was normally distributed. Tukey's Ladder of Powers were therefore applied to successfully transform the BRIEF-2, SCAS and WHODAS 2.0 raw total scores, Next, I calculated *z*-scores for the block design, vocabulary, matrix reasoning and similarities subtests of the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011) individually using the mean and standard deviation of the non-autistic sample, I then calculated a mean of these *z*-scores to create a standardised raw metric of general cognitive ability. The total raw scores from each of the self-reported questionnaires were centred and scaled such that they had a mean of zero and a standard deviation of one. Standardised IQ scores have accounted for age within the scores, raw scores from each task were not used because they were on different scales and so the generation of *z*-scores from the general cognitive assessments enabled age to be included as an independent measure in the regression analyses.

Analysis plan

First, I assessed the convergent validity of the MMEA. To do this I conducted Pearson product-moment correlations between performance on the NIH-EXAMINER and the MMEA in the autistic and non-autistic samples I also applied Fisher's *r*-to-*z* transformations and tested whether the coefficients were comparable across groups. I then conducted a 2 x 2 factorial ANOVA to assess whether there was a main effect of group (autistic, not autistic) or task type (MMEA, NIH-EXAMINER) on standardised performance scores and, in particular, to assess if there was an interaction between group and task type.

Second, I tested whether there were group differences on each of the self-reported questionnaires by conducting one-tailed independent sample *t*-tests. I constructed a multiple linear regression model with stepwise entry for each of the self-reported questionnaire measures to address my primary aim: to determine the extent to which individual differences in age, general cognitive ability, performance on the MMEA and on the NIH-EXAMINER were related to each of the self-reported measures of everyday adjustment (i.e., the BRIEF-2, PedsQL 4.0, SCAS, WHODAS 2.0). I conducted multiple regression models first by entering only performance on the MMEA or the NIH-EXAMINER in order to determine if performance on each assessment related to the measures of everyday adjustment. Then, I conducted a model where performance on both assessments were entered stepwise, to determine if performance on both assessments explained further unique variance in functional adjustment when consider simultaneously.

Convergent validity of the MMEA

There was a significant relationship between the global executive composite generated from the NIH-EXAMINER battery and the composite score created on the MMEA among non-autistic participants, $r(29) = .6$, $p < .001$, which provided convergent validity that the MMEA and NIH-EXAMINER were assessing the same underlying construct, at least among non-autistic participants. There was, however, no relationship between performance on the two tasks among non-autistic participants, $r(31) = -.01$, $p = 0.98$, suggesting that EF ability was not being equally captured by these tasks, among autistic participants. See Figure 17 for each of the correlation coefficients.

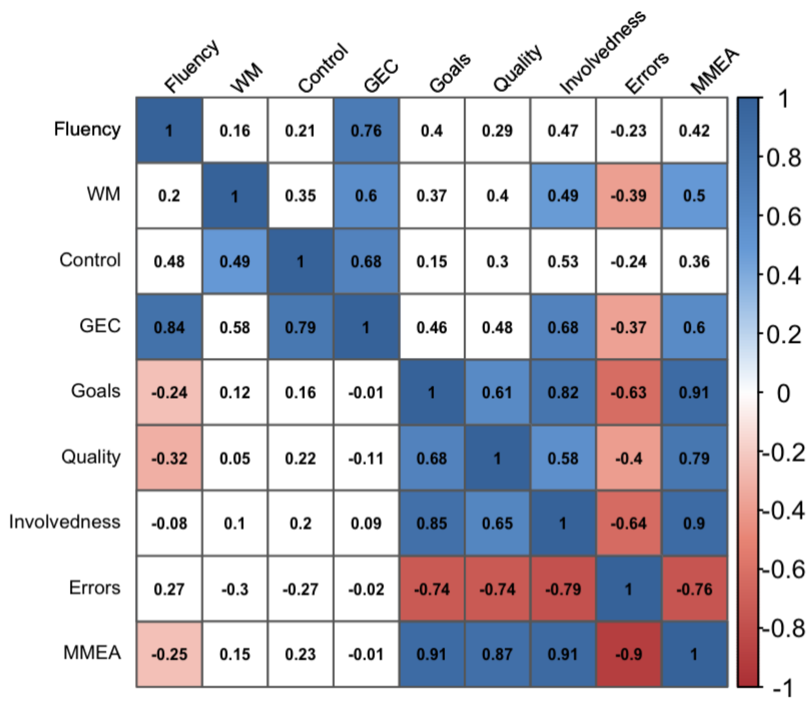


Figure 17. Pearson product-moment correlations between each of the factor scores from the NIH-EXAMINER battery (fluency, working memory, cognitive control and global executive composite) and each of the dependent variables from the MMEA (proportion of completed goals, quality of goals, goal involvedness, executive errors and performance composite). Correlations were calculated for the autistic and non-autistic samples separately, correlation coefficients for the autistic sample are presented below the horizontal and correlation coefficients for the non-autistic sample are presented above the horizontal. Colour and hue indicate direction and strength of the relationship; coloured boxes indicate a significant correlation.

To test whether there was an interaction between group (autistic, not autistic) and assessment type (NIH-EXAMINER, MMEA), I conducted a 2 x 2 ANOVA with standardised performance scores as the dependent variable. There was a main effect of group, $F(1, 124) = 9.64$, $MSE = 1.29$, $p = .002$, $\eta^2_p = .072$, such that autistic participants underperformed relative to non-autistic participants on the NIH-EXAMINER and the MMEA. There was no main effect of assessment type, $F(1,124) = 0.29$, $MSE = 1.29$, $p = .589$, $\eta^2_p = .002$, and no significant interaction effect, $F(1,124) = 0.28$, $p = .601$, $\eta^2_p = .002$ (see Figure 18).

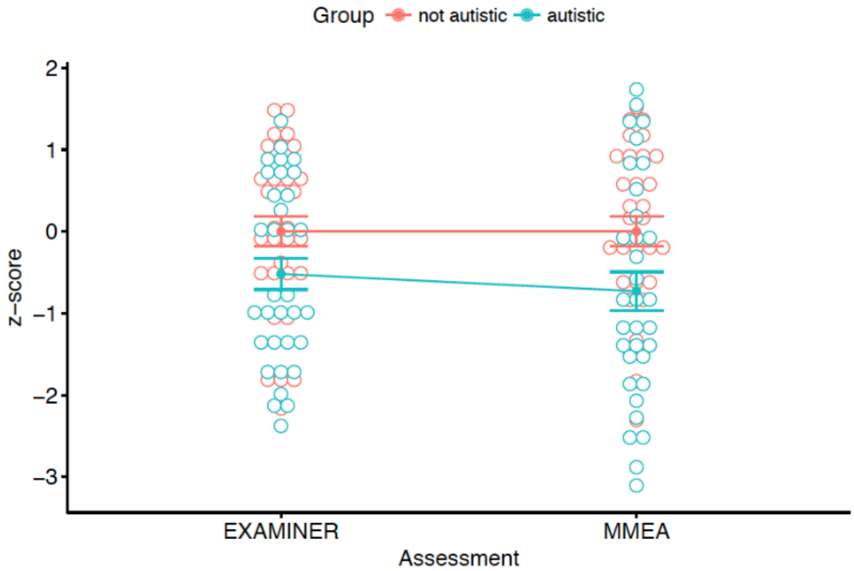


Figure 18. Standardised scores from the Modified Multitasking Evaluation for Adolescents (MMEA) and the National Institute of Health, Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research (NIH-EXAMINER) battery for autistic (n = 33) and non-autistic (n = 31) participants.

Group differences on self-reported questionnaires

Table 13 shows the descriptive statistics for the four self-reported questionnaire measures. There were significant group differences on the raw total scores of the BRIEF-2, $t(62) = 3.56$, $p < .001$, $d = 0.89$, the PedsQL 4.0, $t(62) = 2.32$, $p = .012$, $d = 0.58$, and the WHODAS, $t(62) = 2.96$, $p = .002$, $d = 0.74$, such that autistic participants rated themselves as having more everyday EF difficulties, poorer quality of life and as being more disabled than non-autistic participants. Unexpectedly, there was no group difference on the SCAS, $t(62) = 0.55$, $p = .710$, $d = 0.14$ (see Figure 19).

Table 13.
Descriptive statistics for the self-reported questionnaires

	Non-autistic (n = 31)		Autistic (n = 33)	
	M	(SD)	M	(SD)
	Range		Range	
BRIEF-2	78.81	(15.11)	95.79	(20.98)
	55 - 123		56 - 135	
PedsQL 4.0	74.46	(14.74)	65.90	(14.72)
	45 - 98.33		40.83 - 97.50	
SCAS	31.26	(19.16)	32.30	(15.69)
	6 - 75		5 - 78	
WHODAS 2.0	10.95	(9.99)	17.80	(10.11)
	0 - 40.24		1.04 - 35.04	

Note. BRIEF-2 = Behavior Rating Inventory of Executive Function, second edition (Gioia et al., 2015), PedsQL 4.0 = Pediatric Quality of Life Inventory, fourth edition (Varni et al., 2001), SCAS = Spence Child Anxiety Scale (Spence, 1998), WHODAS 2.0 = World Health Organization Disability Assessment Schedule, second edition (Garin et al., 2010).

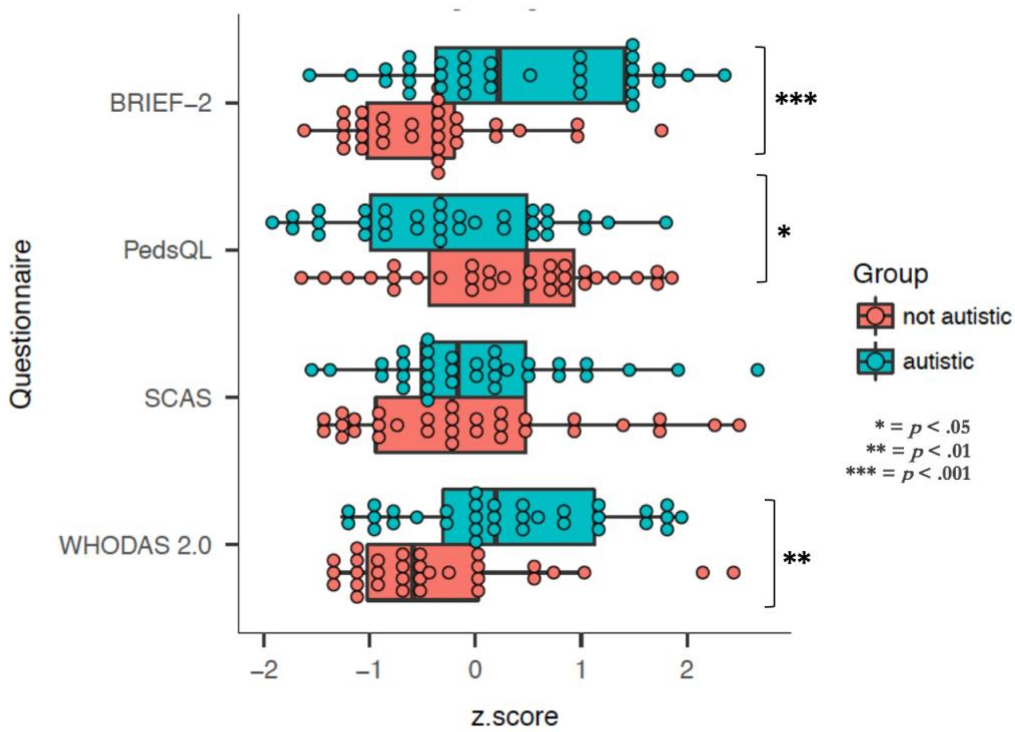


Figure 19. Participant scores on each of the self-reported questionnaire measures for autistic (n=33) and non-autistic (n=31) samples. BRIEF-2 = Behaviour Rating Inventory of Executive Function, second edition, PedsQL = Pediatric Quality of Life Inventory, fourth edition, SCAS = Spence Child Anxiety Scale, WHODAS 2.0 = World Health Organization Disability Assessment Schedule, second edition.

Regression analyses

A series of linear regression models were created to test whether individual differences in general cognitive ability, chronological age, NIH-EXAMINER composite scores and MMEA composite scores predicted individual differences in total raw scores on the BRIEF-2, SCAS, PedsQL 4.0 or the WHODAS 2.0 in the non-autistic or autistic samples separately[10] (see Table 16 for results of such analyses).

In the non-autistic sample chronological age explained approximately 11% of variance in SCAS scores, $F(1, 29) = 4.74$, $p = .038$, *adjusted* $R^2 = .11$. No other models were significant (all *p*s > .05).

In the autistic sample, general cognitive ability and performance on the NIH-EXAMINER made significant unique contributions to explaining the variance on BRIEF-2 scores, $F(3, 29) = 4.44$, $p = .011$, *adjusted* $R^2 = .24$, performance on the MMEA was added to the final model but did not make a significant unique contribution. Performance on the MMEA explained unique variance in trait anxiety, indexed by total raw SCAS scores, $F(1, 31) = 8.14$, $p = .008$, *adjusted* $R^2 = .18$. No other predictors were entered in the final model. Performance on the MMEA explained unique variance in QoL, indexed by total raw scores on the PedsQL, $F(2, 30) = 3.38$, $p = .047$, *adjusted* $R^2 = .13$. Performance on the NIH-EXAMINER was entered into the final model but did not make a significant unique contribution. When considering self-rated disability, as indexed by the WHODAS 2.0, performance on the MMEA was entered into the final model but no significant model could be determined.

[10] Separate regression models were created for autistic and non-autistic participants because I did not have enough statistical power to reliably assess the potential main effect of diagnostic and each of the possible interaction effects.

Table 14.

The final multiple linear regression models when age, general cognitive ability and the NIH-EXAMINER were entered with stepwise entry for each self-reported questionnaire measure, by group

	Not autistic				***Autistic***			
	B	***SE B***	***t***	***Adjust. R^2***	***B***	***SE B***	***t***	***Adjust. R^2***
BRIEF-2				**-**				**0.24***
General cognitive ability	-	-	-		0.42	0.19	**2.21***	
Age	-	-	-		0.13	0.08	1.63	
NIH-EXAMINER	-	-	-		-0.37	0.17	**-2.22***	
SCAS				**0.11***				0.09
General cognitive ability	-	-	-		-0.09	0.17	-0.59	
Age	0.20	0.09	**2.18***		-	-	-	
NIH-EXAMINER	-	-	-		-0.33	0.16	**-2.06***	
PedsQL 4.0				0.06				0.05
General cognitive ability	-	-	-		-	-	-	
Age	0.14	0.08	1.72		-0.14	0.08	-1.69	
NIH-EXAMINER	-	-	-		-	-	-	
WHODAS 2.0				-				-
General cognitive ability	-	-	-		-	-	-	
Age	-	-	-		-	-	-	
NIH-EXAMINER	-	-	-		-	-	-	

Note. BRIEF-2 = the Behaviour Rating Inventory of Executive Function, second edition (Gioia et al., 2015); SCAS = Spence Children's Anxiety Scale (Spence, 1998); PedsQL 4.0 = the core generic scales and the well-being supplement of the Pediatric Quality of Life Inventory (Varni et al., 2001); NIH-EXAMINER = Global executive composite calculated based on performance on the NIH-EXAMINER battery (Kramer et al., 2014).

· *p* < .1, * *p <.05,* ** *p < .01,* *** *p < .001*

Table 15.

The final multiple linear regression models when age, general cognitive ability and the MMEA were entered with stepwise entry for each self-reported questionnaire measure, by group

	Not autistic				***Autistic***			
	B	***SE B***	***t***	***Adjust. R^2***	***B***	***SE B***	***t***	***Adjust. R^2***
BRIEF-2				-				**0.19***
General cognitive ability	-	-	-		-	-	-	
Age	-	-	-		0.15	0.08	1.83·	
MMEA composite	-	-	-		0.21	0.12	1.74·	
SCAS				**0.11***				**0.18****
General cognitive ability	-	-	-		-	-	-	
Age	0.20	0.09	**2.18***		-	-	**-**	
MMEA composite	-	-	-		0.30	0.11	**2.85****	
PedsQL 4.0				0.06·				**0.10***
General cognitive ability	-	-	-		-	-	-	
Age	-0.14	0.08	-1.72·		-	-	-	
MMEA composite	-	-	-		-0.25	0.12	**-2.10***	
WHODAS 2.0				-				0.08·
General cognitive ability	-	-	-		-	-	-	
Age	-	-	-		-	-	-	
MMEA composite	-	-	-		0.22	0.11	1.96·	

Note. BRIEF-2 = the Behaviour Rating Inventory of Executive Function, second edition (Gioia et al., 2015); SCAS = Spence Children's Anxiety Scale (Spence, 1998); PedsQL 4.0 = the core generic scales and the well-being supplement of the Pediatric Quality of Life Inventory (Varni et al., 2001); MMEA = composite performance score from the Modified Multitasking Evaluation for Adolescents (M. Schneider et al., 2016).

· *p* < .1, * *p <.05,* ** *p < .01,* *** *p < .001*

Table 16.

The final multiple linear regression models when age, general cognitive ability, MMEA and NIH-EXAMINER were entered with stepwise entry for each self-reported questionnaire measure, by group

	Not autistic				***Autistic***			
	B	***SE B***	***t***	***Adjust. R^2***	***B***	***SE B***	***t***	***Adjust. R^2***
BRIEF-2				-				**0.24***
General cognitive ability	-	-	-		0.41	0.19	**2.14***	
Age	-	-	-		-	-	-	
NIH-EXAMINER	-	-	-		-0.40	0.16	**-2.42***	
MMEA	-	-	-		0.20	0.12	1.69	
SCAS				**0.11***				**0.18****
General cognitive ability	-	-	-		-	-	-	
Age	0.20	0.09	**2.18***		-	-	-	
NIH-EXAMINER	-	-	-		-	-	-	
MMEA	-	-	-		0.30	0.11	**2.85****	
PedsQL 4.0				0.06·				**0.13***
General cognitive ability	-	-	-		-	-	-	
Age	-0.14	0.08	1.72·		-	-	-	
NIH-EXAMINER	-	-	-		0.21	0.14	1.48	
MMEA	-	-	-		-0.25	0.12	**-2.13***	
WHODAS 2.0				-				0.08
General cognitive ability	-	-	-		-	-	-	
Age	-	-	-		-	-	-	
NIH-EXAMINER	-	-	-		-	-	-	
MMEA	-	-	-		0.22	0.11	1.96	

Note. BRIEF-2 = the Behaviour Rating Inventory of Executive Function, second edition (Gioia et al., 2015); SCAS = Spence Children's Anxiety Scale (Spence, 1998); PedsQL 4.0 = the core generic scales and the well-being supplement of the Pediatric Quality of Life Inventory (Varni et al., 2001) NIH-EXAMINER = Global executive composite calculated based on performance on the NIH-EXAMINER battery (Kramer et al., 2014); MMEA = composite performance score from the Modified Multitasking Evaluation for Adolescents (M. Schneider et al., 2016).

· $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

This study investigated the convergent validity of a representatively-design EF task with a more traditional approach to EF assessment and showed that, for non-autistic participants, performance on both tasks had a strong positive relationship. There was however no relationship between performance on both tasks for autistic participants. One possibility is that while both assessments appear to be measuring the same underlying cognitive construct in non-autistic participants, namely EF, the composite measure generated also captures individual differences introduced by the non-executive components of the MMEA (e.g., social, sensory and motor) and the lack of relationship with performance on the NIH-EXAMINER among autistic participants is because these factors impact performance more for the autistic sample. Alternatively, it might be that while neuropsychological tasks such as those included in the NIH-EXAMINER capture variability in EF in non-autistic participants, they may be measuring variability in other abilities that were not the target of the assessment such as social desirability, motivation to persist with monotonous tasks or the ability to infer implicit information about the tasks (White, 2013).

The possibility that group differences on classic EF paradigms might reflect differences in non-executive abilities was, in part, the motivation to assess the EF ability of autistic adolescents on a function-led assessment such as the MMEA. To assess whether the highly-structured nature of typical EF assessment were leading to an understatement of EF difficulties in autism, thus presenting an autism-specific threat to the ecological validity of EF assessment, the present study assessed whether assessment type interacted with group. While autistic participants struggled with EF in general, they did not appear to struggle to a disproportionate extent on one assessment (which would have been evidenced by a significant group x assessment type interaction), contrary to my hypothesis. This is not easily reconciled with the recurrent finding in the literature that autistic people tend to have more difficulties on less-structured assessments, nor does it sit neatly alongside the finding that autistic participants demonstrated less competence on the overall task.

I showed that autistic participants reported more difficulties with respect to their everyday EF ability when compared with non-autistic participants. This was the first study, to my knowledge, where the self-report BRIEF questionnaire was administered to autistic adolescents but it is nevertheless consistent with numerous studies demonstrating adolescents EF difficulties, measured by the *parent-report* BRIEF (Pugliese et al., 2015, 2016; Rosenthal et al., 2013). The finding that autistic adolescents rate their quality of life as lower than their non-autistic peers is concerning and is in keeping with previously published reports (Kamp-Becker, Schröder, Remschmidt, Bachmann, & Schroder, 2010; Shipman et al., 2011). Autistic participants also rated themselves to be more disabled, on average, than the non-autistic participants, just like adult participants in a recent study to use the WHODAS 2.0 (L. Schmidt et al., 2015), There were no differences, however, in self-report ratings of trait anxiety across groups, which was surprising since many studies have shown elevated rates of anxiety in autistic samples (see White, Oswald, Ollendick, & Scahill, 2009, for review), however the non-autistic sample (as well as the autistic sample) in the current analysis had elevated rates of anxiety relative to a normative population sample. Further, group differences between autistic and non-autistic young people on the SCAS appear to be more pronounced in parent- rather than self-reported versions of the questionnaire (Boulter et al., 2014).

Veridicality should be reflected in significant relationships between laboratory-based assessments and measures of everyday functioning. Here, I showed evidence of such veridicality in the autistic sample, by showing that performance on the NIH-EXAMINER, alongside general cognitive ability, was related to self-reported everyday EF ability which demonstrated the generalisability of the classic paradigms of EF assessment when assessed against the BRIEF-2 – consistent with the finding that older autistic adults self-reported BRIEF scores correlated significantly with their scores on a performance-based measure of

EF (the Tower of London task; Davids et al., 2016). Performance on the MMEA was related to participants' self-reported anxiety, which is in keeping with a previously reported relationship between EF and anxiety in autistic adolescents, albeit when neuropsychological measures were used (Hollocks et al., 2014). Similarly, performance on the MMEA was associated with participants' quality of life. It is noteworthy that, in both cases, individual differences in general cognitive ability were not entered in the final regression models. The lack of a contribution of general cognitive ability here suggests that despite IQ being an important prognostic predictor of outcome in autistic people more broadly (e.g., Howlin et al., 2004), individual differences in EF ability might be a more important predictor of aspects of real-life functioning, at least for cognitively able autistic individuals, something that has been shown in typical development (Moffitt et al., 2011) and in autism (Kenny et al., 2018). The relationship presented in the current study is, however, cross-sectional and needs to be tested longitudinally to determine the relative predictive power of specific executive abilities, as assessed both by laboratory-based and real-world tasks, and general cognitive ability over time.

Neither the MMEA nor the NIH-EXAMINER were related to self-perceived disability in our samples. This finding is surprising because, although this construct has not previously been assessed among autistic adolescents, associations between measures of EF and the closely related concept of adaptive functioning have been documented both concurrently (Happé, Booth, et al., 2006; Panerai et al., 2014) and longitudinally (Kenny et al., 2018; Visser, Berger, Lantman-De Valk, Prins, & Teunisse, 2015). Given the transdiagnostic applicability of the WHODAS 2.0, it is possible that it lacks specificity about the types of disabling consequences of an autism diagnosis in particular.

Despite this study providing evidence for the generalisability of performance on these measures to real-world realities for autistic participants, we failed to demonstrate such evidence for the non-autistic sample: neither performance on the MMEA nor on the NIH-EXAMINER explained unique variance in any of the self-reported measures. This is consistent with previous investigations that have shown weak or no relationships between EF tasks and self-reported measures of self-control (Duckworth & Kern, 2011), inhibition (Cyders & Coskunpinar, 2011), impulsivity (King, Patock-Peckham, Dager, Thimm, & Gates, 2014) and self-regulation (Eisenberg et al., 2017) in typical adolescents and adults. This raises questions regarding the generalisability of these EF measures to other real-world measures, at least as assessed in non-autistic adolescents.

Understanding whether the systematically-designed or the representatively-designed task, assessed by the current study, demonstrated comparatively greater generalisability depends on the metric of real-world functional adjustment against which it is assessed. Neuropsychological measures, such as those included in the NIH-EXAMINER, are related to participants' own reports of their everyday EF ability. This provides evidence for the convergent validity of these tasks, i.e., they both measure the same construct (although see Toplak, West, & Stanovich, 2013, for an argument to the contrary) and it provides evidence that this task has veridicality. In fact, the lack of relationship between performance on the MMEA and BRIEF raises some questions, particularly if one considers that the BRIEF measures everyday EF ability which should be more closely related to an everyday-type task than to artificial laboratory tasks. If, however we consider the more far-reaching consequences EF difficulties can have, such as lower quality of life and greater levels of experienced anxiety, then the MMEA appears to have greater veridicality than does the NIH-EXAMINER. Although, by that logic, one should conceivably expect that variance in EF abilities should explain variance in autistic peoples' perceptions of the extent to which they are disabled and yet, no such relationship was found.

The real-world measures selected here were questionnaires, a decision that was taken with consideration for the feasibility of the current study but the generalisability of results from EF assessments need to be assessed against observations of participants in the real-world, either by trained researchers or reported on by participants in real-time using experience sampling methodologies or compared with metrics of functioning such as educational or occupational status or performance.

In summary, the current study demonstrated that autistic participants had more difficulties on most of the measures selected to assess the veridicality of EF assessment against. On the one hand, performance on the construct-driven paradigms, borrowed from the disciplines of experimental and neuropsychology, does generalise to participants' self-report about everyday EF ability. On the other hand, performance on the function-driven approach, to assess cognitive abilities in context resembling the context over which we aim to generalise the findings, generalises to participants' self-reported anxiety and quality of life. An important finding from the present analysis is that neither task type is related to the extent to which participants report feeling disabled. In order to understand how autistic adolescents, conceptualise their EF abilities and disabilities we need to ask them directly how they think and feel about completing executively demanding tasks and whether they perceive executive abilities to have an impact on their lives. Furthermore, our ability to devise function-led approaches to assessing the type of EF difficulties that autistic people report themselves as having on self-report questionnaire measures, but which remain relative elusive to measurement in EF tasks, we must ask autistic young people what executive related difficulties they consider to impact on their lives, if any, how these are intertwined with other non-executive skills and whether they believe their being autistic is implicated in these relationships.

Chapter 5

The reality of executive function difficulties for autistic adolescents: personal and parental perspectives

Introduction

So far, I have shown that autistic adolescents displayed more difficulties, on average, than non-autistic adolescents on a battery of systematically-designed executive function (EF) tasks (Chapter 2), on a representatively-designed EF task (Chapter 3) and that performance on each of these tasks was related to at least one metric of self-reported everyday adjustment (Chapter 4). Nevertheless, our ability to know whether these difficulties are relevant to the lives of young autistic people remains limited by the methodologies hitherto applied in EF research. Specifically, the voices of young autistic people are completely absent from the literature pertaining to EF in autism. Therefore, my final empirical chapter sought to address this issue by eliciting the views and perspectives of young autistic people and their caregivers about their own or their children's EF.

The autistic 'voice' is all too often absent from autism research (Milton, 2014). In relation to EF in autism, this absence is problematic for three reasons. First, studies typically employ measures of EF designed by non-autistic researchers and more typically still by researchers who are neither autistic nor researchers whose experience relates to autism. Consequently, we can test only whether autistic people have EF difficulties similar to those found in other clinical populations and are precluded from identifying instantiations of EF difficulties that are specific to autism. Furthermore, imposing the narrow, supposedly objective standards of functioning developed by non-autistic researchers can serve to mischaracterise and often underestimate the abilities of autistic people (Ruble & Dalrymple, 1996). Second, performance-based and questionnaire-based measures of EF rarely, if ever, tap participant perceptions of their own executive abilities or the impact EF difficulties have upon their lives. Uncovering whether the group differences measured in research are perceptible to participants or understanding the salience of these differences to an individual autistic person is therefore not currently possible. Third, classifying somebody as having an EF difficulty based on their behaviour during a task on a single occasion or on their rating of how frequently they have experienced specific difficulties is potentially problematic. It may conflate a general EF difficulty with the effects of contextual or antecedent factors that result in a person's ability to perform executively demanding activities being compromised, for example, when their senses are overloaded or their motivation to complete the task is low.

While questionnaire-based measures, which facilitate subjective reporting, are common in the measurement of EF in adolescents on the autism spectrum, they have almost exclusively involved an informant rating on an autistic person's executive ability and therefore, critically, do not capture subjective first-person accounts (Rosenthal et al., 2013; van den Bergh, Scheeren, Begeer, Koot, & Geurts, 2014). One often-cited reason for not asking autistic young people themselves is that they might not show the requisite insight into their own behaviour in order to provide valid and reliable data (U. Frith & Happé, 1999; Zahavi, 2010). There are, however, a few exceptions to this pattern. In one study, adolescent autistic males rated themselves on the Dysexecutive Questionnaire as having EF difficulties that were, on average, as severe as groups with traumatic brain injuries (Cederlund, Hagberg, & Gillberg, 2010). Furthermore, autistic participants in early (Dijkhuis et al., 2017), middle (Wallace et al., 2016) and late (Davids et al., 2016) adulthood have rated themselves as having more EF difficulties than non-autistic participants on the adult version of the Behaviour Rating Inventory of Executive Function (BRIEF-A). Self-reported BRIEF-A has been shown to be related to subjective ratings of quality of life (Dijkhuis et al., 2017), mental health symptomatology (Wallace et al., 2016) and performance on a classic laboratory measure of EF (the Tower of London task) as well as proxy-reported BRIEF-A (Davids et al., 2016). Taken together, these findings demonstrate that autistic people do not necessarily lack insight into their own EF difficulties and so they should not be considered unreliable.

Qualitative research has traditionally been underutilised in autism research, due to a lack of methodological expertise among researchers, a perceived lack of need and a supposed lack of rigour (Bölte, 2014). This is changing. Researchers are becoming persuaded by the critical importance of engaging, rather than excluding the autistic voice from research (Fletcher-Watson et al., 2018; Gillespie-Lynch, Kapp, Brooks, Pickens, & Schwartzman, 2017; Pellicano, Ne'eman, & Stears, 2011) and because of the proliferation of rigorous qualitative autism research (e.g., R. S. Smith & Sharp, 2013). Below, I provide a brief overview of qualitative research that has examined the lived experiences of autistic adolescents and their parents, with a specific focus on descriptions related to everyday EF abilities.

Qualitative research focusing on autistic adolescents

A meta-synthesis, the qualitative equivalent of a meta-analysis, of autistic young peoples' experiences in mainstream school settings found the most pertinent themes to emerge were that autistic adolescents were acutely aware of being different and had difficult sensory and interpersonal experiences (E. I. Williams, Gleeson, & Jones, 2017). EF-related difficulties were notable by their absence from this synthesis. Another meta-synthesis, with the broader aim of understanding the experiences of autistic young people and adults from their own perspective, did however yield descriptions of EF-related difficulties under the themes of school experiences and work-related factors (DePape & Lindsay, 2016). For instance, in relation to managing workloads and the role rigid routines play in alleviating stress that results from executively demanding tasks.

One study, in which nine autistic students were interviewed about being educated in an inclusive setting, revealed issues with the curriculum, including a heavy or unmanageable workload, self-managing the timescale for assignments and tight deadlines (Saggers, Hwang, & Mercer, 2011). Similarly, according to parent- and self-report, autistic children's inability to adapt flexibly to changing routines can bring with it challenges for the whole family (Ludlow, Skelly, & Rohleder, 2012) as well as at school (Humphrey & Lewis, 2008). This finding is consistent with the reports of two autistic adolescents and their mothers, who describe similar school-related stresses and mention that the rigid adherence to routines offsets some of the stress caused by the ever-changing demands of school life (Carrington & Graham, 2001).

Another important consideration for the assessment of EF in autism that was apparent was that some autistic people reported that their performance on an activity is greatly influenced by their level of interest in that activity. Autistic young people have reported being bored or disinterested in their school experience and reported only applying themselves to subjects and experiences that captivate their interests (Marks, Schrader, Longaker, & Levine, 2000). This has important implications for the assessment of EF capability in autism: if participants are not asked about their disposition toward a task and the context in which they are assessed, we cannot know how much of the variance in their performance is attributable to EF difficulties and how much is attributable to a lack of interest.

Although not a description of EF per se, autistic young people often report having difficulties understanding rapidly-delivered task instructions (Preece & Jordan, 2010), particularly when delivered in the context of rich, often overstimulating, sensory environments (R. S. P. Jones, Quigney, & Huws, 2003). Such difficulties understanding task instructions can extend to broader contexts, where rules may change or be applied flexibly, such as the rules governing school life (Penney, 2013). Consequently, we must remain open to the possibility that underachieving on measures of EF may, at least to an extent, result from difficulties understanding explicit instructions or to inferring the instructional content that is typically implicitly inferred.

Autistic people's social difficulties are often reported as trumping non-social difficulties (Ludlow et al., 2012), but the increased need for independence that comes with advancing age often elicits references to EF abilities. In particular, realisations that these abilities might be incommensurate with developmental expectations with respect to managing school assignments (Connor, 2000), self-care (Galpin et al., 2017) and transitioning from school to post-secondary settings (Browning, Osborne, & Reed, 2009; Cribb, Kenny, & Pellicano, 2018). For instance, autistic adolescents who were approaching or had recently completed secondary education in Western Australia reported having difficulties with future-oriented thinking and managing the increasing expectations for independence placed on them as they transition up into adult roles (Cribb et al., 2018). Similarly, in another study, adolescents approaching the end of secondary education in the UK reported being worried about whether their cognitive and behavioural capacities would allow them to succeed in their future lives (Browning et al., 2009).

Although no study, to my knowledge, has directly asked autistic people to discuss how EF affects their life chances, there are many instances when autistic people have spoken about their autistic identity, which have implications for understanding the ways autistic people conceptualise their own thoughts and behaviours. Some interpret their perceived difference and their being autistic positively, accepting it as an essential part of who they are as a person, to be celebrated, others appear somewhat ambivalent, and others still describe it with despondency, anger and frustration (Cribb et al., 2018; Hebron & Humphrey, 2014; Humphrey & Lewis, 2008; McLaughlin & Rafferty, 2014). Despite the awareness that being autistic presents challenges, including with organisation and planning, many autistic young people distance themselves from autistic and disabled identities, feeling that such labels serve to negatively affect how others perceive them (Baines, 2012; Hull et al., 2017). The provision of appropriate support is repeatedly cited as critical to counter the negative consequences often associated with being on the autism spectrum, such as, for example, inflexibility (Hay & Winn, 2005) and subsequently precipitating mental health difficulties (Crane et al., 2018; Hebron & Humphrey, 2014). As well as a reluctance to identify themselves as somebody who would benefit from support, poor organisational skills can themselves act as a barrier to accessing support (Nicolaidis, Kripke, & Raymaker, 2014). Understanding the nature of autistic young people's difficulties *and* how these difficulties are understood and conceived of in relation to a person's constructed sense of self are both necessary steps to identify how to effectively and sensitively provide support to those with poor executive abilities.

The current study

The current study examined directly the views and experiences of autistic adolescents and their parents towards their executive skills. It also sought to understand the perceived consequences of any EF difficulties and whether they conceive of their EF difficulties as related to being autistic.

Method

Participants

Twelve autistic adolescents (11 male, 1 female; all participants are referred to with male pronouns to maintain participant anonymity), aged between 12 years and 19 years, and seven of their parents (all mothers, one parent spoke about her two children) were interviewed. One additional parent participated whose son decided not to be involved in the study. Seven of these participants completed the assessments detailed in Chapters 2, 3 and 4. Five participants were recruited for this study to increase the sample size. Participants all had IQ scores within at least the typical range, as measured by the Wechsler Abbreviated Scale of Intelligence, second edition (Wechsler, 2011). All participants had received an independent clinical diagnosis of autism, which was confirmed by a parent-reported, Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012) *t*-score of at least 60, indicative of an autism spectrum disorder. Eight participants were attending a mainstream secondary school, one participant was attending an autism-specific special school, one was in a Further Education College, one participant was home-schooled, and one participant was not accessing any form of education at the time of the interview. Four participants had at least one co-occurring condition, as shown alongside participant background characteristics in Table 17.

Table 17.
Background information relating to the participants who were interviewed

Participant	Gender	Age	Co-occurring disabilities	IQ classification	Autism severity	School setting
1	Female	16	DCD, PDA	Low average	Severe	Not in education
2	Male	12	None	High average	Severe	Mainstream[a]
3	Male	13	ADHD, DCD, Tourette's	High average	Moderate	Special[b]
4	Male	14	None	Average	Severe	Mainstream[c]
5	Male	18	None	Borderline	Mild	Mainstream[d]
6	Male	14	None	Low average	Moderate	Mainstream[d]
7	Male	19	ADHD, DCD, Dyslexia	Average	Severe	FE college
8	Male	14	None	Average	Moderate	Mainstream
9	Male	13	ADHD	Average	Severe	Mainstream[a]
10	Male	12	None	Average	Severe	Home educated
11	Male	14	None	High average	Moderate	Mainstream[d]
12	Male	15	None	Low average	Severe	Mainstream[d]

Note. ADHD = Attention Deficit Hyperactivity Disorder, DCD = Developmental Coordination Disorder, FE = Further education, PDA = Pathological Demand Avoidance. IQ classification = classification based on the full-scale intellectual quotient (IQ) from the Wechsler Abbreviated Scale of Intelligence, second edition (WASI-II; Wechsler, 2011); extremely low = IQ < 70, Borderline = 71 < IQ < 80, Low average = 81 < IQ < 90, Average = 91 < IQ < 110, High average = 111 < IQ < 120, Superior = 121 < IQ < 130, Very superior = 131 < IQ < 140. Autism severity = classification based on the t-score from the Social Responsiveness Scale, second edition (SRS-2; Constantino & Gruber, 2012); Mild = 60 < t < 65, Moderate = 66 < t < 75, Severe = t >76.
[a] Mainstream school with a support base that is not autism-specific; [b] An autism-specific special school; [c] A mainstream school with a support base that is autism-specific; [d] A mainstream classroom with a learning support assistant

Measures

Semi-structured interviews. Autistic adolescents and parents of autistic adolescents were interviewed individually using semi-structured interviews. The same interview schedule was used with the young people and their parents. They were asked a number of open-ended questions about their perceptions of the participants' EF abilities, with a particular focus on their higher order abilities, such as the ability to manage their time, to multitask, to retain information and to adapt flexibly to changes in task demands.

Questions were sufficiently broad to allow participants to introduce the specific experiences they felt were relevant to the topic and to elaborate on themes important to them; probe questions were asked when necessary. The young autistic people and their parents were first asked a general question about how they or their child were currently getting on, with respect to school and home life. This question was to provide background information to the interviewer and to encourage participants to reflect on their current abilities, although reflections on their abilities at younger ages were freely explored if they were mentioned. A critical incident technique (Flanagan, 1954) was used, whereby participants were asked to recollect specific situations where they (or their child) excelled at or had difficulties employing EF abilities. For example, I asked questions like, "can you tell me about a time where you/your child were really proud of a plan you/they made?" and "can you think of a time when you/your child found it hard to multitask?" (for an example application of this technique, see Dickie, Baranek, Schultz, Watson, & McComish, 2009). Interviews were recorded with participants' prior consent and professionally transcribed verbatim. One participant opted not to have their interview recorded and so detailed notes were taken immediately after the interview was completed.

Analytic technique, epistemology and positionality

Transcripts were analysed using Thematic Analysis, following the six steps outlined by Braun and Clarke (2006), namely, (i) familiarisation with the data, (ii) creating initial codes, (iii) searching for themes, (iv) reviewing themes, (v) naming themes and (vi) producing the report. The transcripts were analysed from an inductive (bottom-up) perspective where themes were created within a 'contextualist' framework of critical realism (Bhaskar, 1978). Critical realism lies between a realist ontology, which asserts that only one objective state of reality exists, irrespective of the researcher's or participant's views on it (Jenkins, 2010) and a relativist ontology, which asserts that reality is entirely constructed through historical, political and social interchanges (Schwandt, 1994). Realist ontologies are associated with a positivist epistemology, which aims to take reliable and valid measures of objective phenomenon and relativist ontologies are associated with a constructionist epistemology, which aims to uncover created realities, which cannot be separated from the person who created, described and lived that reality. Critical realism serves as both an ontological and epistemological position that has three overlapping domains of reality, including the empirical, the actual and the real (Fletcher, 2017). That is, critical realism acknowledges that we all have subjective experiences (the empirical), that an objective reality exists outside of our experience (the actual) and that causal mechanisms lie between and within these domains (the real). Critical realism is therefore suited to the current study because it allows for EF difficulties existing as a result of brain-based differences, irrespective of a participant's views on them (the actual reality) but also allows space to tap into participants' subjective experiences of these difficulties (the empirical reality) and facilitates discussion of how these might interact (the real reality).

Some researchers assess the inter-rater reliability between the codes generated in analysing the themes that emerge in qualitative analysis. This practice serves the objectives of a realist epistemology and positivist ontology, which assert that one objective reality exists in the data and by extension there should be reliable agreement between multiple people analysing the same data. In the current study, I was approaching analysis from a critical-realist perspective, whereby there is no assumption that a single objective reality exists within the data presented, instead, the themes created reflect my personal conceptualisation of the themes created between me and the participants. It was appropriate or consistent with a critical realist approach to assess if the codes and themes that I created were reliable with those generated by another researcher, whose positionality and proximity to the data and to the conceptualisation of EF and issues with its measurement would have differed. An associated limitation of this approach is that this analysis cannot be generalised beyond the context of this study and researchers with a more positivist inclination may find the claims presented as lacking veracity.

Interviews for the young people and their parents were considered together because rather than seeking to compare and contrast the themes emerging from these two groups, something that qualitative methodologies are poorly suited to do (Pope & Mays, 1995), I sought to understand the experiences of young autistic people both from first- and third-person perspectives. Triangulation, including seeking the perspectives of two informants, is an often employed credibility check in qualitative research (Elliott, Fischer, & Rennie, 1999; Patton, 1999).

Elliott, Fischer and Rennie (1999, p. 221) emphasise the importance of "owning one's perspective" in qualitative research. This includes the author of a piece of qualitative research specifying their theoretical orientations and anticipations for the resultant research. Explicitly communicating one's positionality, according to Elliot et al. (1999), allows the researcher to be aware of the impact their views have on constructing meaning in the research as well as helping readers to interpret the data and the researchers interpretation of the data. In qualitative, or indeed in quantitative research, it is neither feasible nor always necessarily desirable for the researcher to remain wholly objective (Schutz, 1994). To that end, researchers often engage in a practice known as reflexive bracketing to minimise their role in shaping the emergent themes and to avoid stifling the participant voice (Ahern, 1999).

I will therefore outline my own position with respect to understanding the EF experiences of autistic adolescents and their parents. I am neither autistic nor am I a parent and so I will necessarily interpret the interviews in this study from an outsider perspective and may misappraise them accordingly. Relatedly, my experience of adolescence was not that of an adolescent who struggled with executive abilities and I did or do not view myself as owning any disabled identity so my interpretation in these respects will likely differ from those who have lived these experiences. My interest in EF abilities in autism was borne out of my previous experiences as an autism researcher with interests in understanding the non-social difficulties autistic people regularly refer to as affecting their life chances. I ascribe to the neurodiversity paradigm (Singer, 1999) and believe autism reflects a natural variation in human neurology, albeit within a social relational model of disability (i.e., disability results from the interplay between an individual's neurodivergence and the societal response to that divergence; Thomas, 2004).

Procedure

Ethical approval for this study was granted through the postgraduate student ethics procedure at UCL Institute of Education. All participants provided written informed consent prior to participating in this study. Participants completed individual semi-structured interviews face-to-face in a quiet room at the UCL Institute of Education (young people, $n = 6$; parent, $n = 1$) or in their own home, or over the phone (young people, $n = 5$, parent, $n = 7$).

Results

Four major themes (presented in **bold**) and up to four subthemes (presented in *italics*) were created from the data. These are displayed in Figure 20 and discussed below.

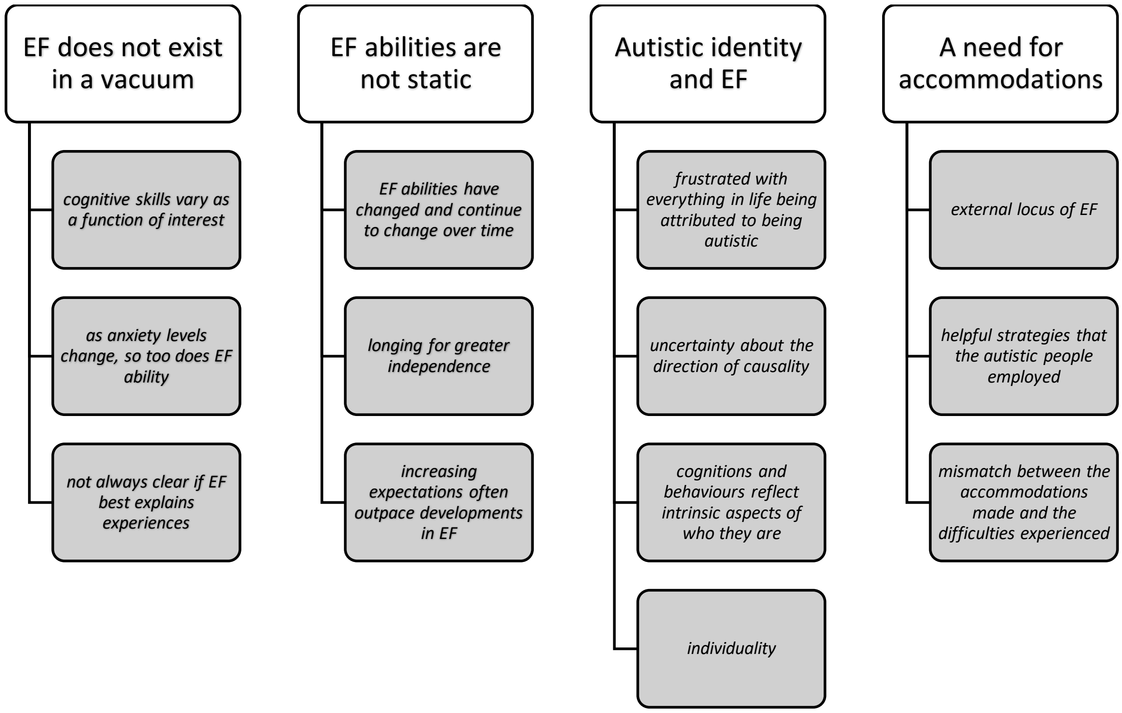


Figure 20. The four major themes and the related subthemes that were identified from the interviews with young autistic people and their parents.

EF does not exist in a vacuum

A central theme identified in the data was that EF is dependent on contextual and motivational factors and so is not fixed, many participants referred to their ability to deploy their *cognitive skills varying as a function of their interest* in the subject at hand or their motivation to do so at a particular moment in time. One young man acknowledged, "if it's something which I really don't like, I find that I'm often a lot slower than if it's something which I really do like" and another reported, "if it's something in which I'm extremely focussed on in terms of interest … then I plan ahead of time quite well". When speaking about her son's performance in school activities, one mother emphasised that underperformance was not always necessarily related to her child's ability but, rather, it was often related to his motivation: "to be honest, a lot of it is he's not really interested, he doesn't want to do it, and so it's not that he can't, it's that he doesn't want to". While one young man acknowledged that executively demanding activities like multitasking were always anxiety provoking, he suggested that if the activity was in the service of a special interest, the trade-off could be worth it:

> If I'm extremely energetic about a certain activity that just, again, so happens to correspond to my interests, then I'll multi-task about those things, if I'm extremely energetic about them. But general multi-tasking, overall on the whole, can sometimes make me feel a bit of stress if, you know, I have to process more than one thing (Participant 4).

One participant referred to their ability to selectively control their cognitive resources, even in situations where they were not especially aligned with their interests: "Well I just force my brain into remembering it, even though it may be boring." This sentiment, however, was not universal among participants.

Another common experience was that *as anxiety levels change, so too does EF ability*. For instance, one mother said of her 19-year-old son:

> I think also, the autism definitely, the Asperger's syndrome, produces this huge anxiety all the time, and when you're anxious, planning becomes much more painful because you're afraid all the time you're going to fail and that you've got it wrong, or that you're not going to be able to do it. So, you know, to me, that's a fairly toxic combination, actually (Parent of Participant 7).

This view highlighted a commonly-held experience where the ability to engage in EF such as planning can be completely eclipsed by overwhelming anxiety, a sentiment reinforced by the participant himself, who acknowledges that his mood often impacted on his cognitive functioning:

> I normally just, I want to stay at home, but sometimes it's difficult to plan ahead. Because you don't know how you're going to feel every day. So, some days you just don't want to get out of bed. So, if you don't want to get out of bed then there's no point in getting out of bed (Participant 1).

A parent who highlighted the importance of her son's mental state in his ability to be flexible echoed this view: "It's like, I think sometimes when his mental state is in a good place, I think he might be able to manage slightly. But when it is not, then it is mayhem." This was confirmed by the young person who, speaking of his inability to attend school, which his school often interpreted as a sign of inflexibility, "Well, I don't really know, it's just I get built up in my head and that stopped me from physically actually being able to get there [school]."

By contrast, another parent believed that her son often functioned well in school but said this was not necessarily because he was intrinsically motivated ("he doesn't care whether he does it well or not, it's not about the learning for him") or because it came easily to him ("keeping organised is not something he does very well") but rather "because he's afraid of detention." According to this participant, her son was highly motivated to behave in certain ways, including keeping himself organised, to avoid detention, which serves to remind us that the relationship between motivation and EF is not straightforward, sometimes EF is detrimental to performance on executive-demanding activities and sometimes it elevates performance, although according to some accounts "this comes at a huge cost".

An important stipulation, made clear by many of the participants in the current study, was that it was *not always clear to them whether EF best explained their experiences*. One recurring point was that participants found it difficult to follow the instructions they were given or to uncover the often, implicit goals that life presented them with. When asked about achieving goals and completing tasks participants suggested that the instructions were insufficiently detailed ("what I've found is oftentimes that instructions aren't detailed enough"), explicit ("because some won't be as like, as direct as I am) or clear ("teachers can be a bit unclear with their explanation at times"). Complex instructions were often difficult for the young people in the study, who found it challenging to retain ("I know she's explained it, but I can't remember. When I come to do it, I can't remember what it was she said to me that I had to do"), execute ("I have to give him one instruction at a time otherwise he just panics and freezes") and avoid the overly literal interpretation of ("because he heard the instruction, he did what he was asked") task instructions.

EF abilities are not static

Both the young autistic people and their parents acknowledged that their *EF abilities have changed and continue to change over time.* For instance, when asked about specific executive abilities, participants felt "I found it hard at first but now I am okay with it", "that has become easier actually" and "he's really getting a lot better on the organisation front which is much better than he used to be, which is great." The development in EF abilities was often raised in the context of participants *longing for greater freedom*. For instance, when asked about planning their own days, one participant said: "I think the reason why I'm not able to do much of the planning myself is because I'm not old enough to drive. Things like that." Another, when asked about the allocation of chores at home reported, "I do sometimes wish that they would leave me to make my own sort of decisions for that, because it's come to be a repetitive routine."

While parents tended to laud their children's increased desire for independence they were also acutely aware that with increasing age, the societal expectations placed on their children continue to change and parents often worried that the *increasing expectations outpaced their developments in EF.* For example, one parent said of her child:

> When I look at him and I think he's such hard work, I have to remind myself that he's come such a long way, and it's a developmental delay. That's what they always say, isn't it, about autism? I think if it's a four-year delay, he's pretty good for a 15-year-old, he's just not good yet for a 19-year-old (Parent of Participant 7).

Another, worried about how her child would cope in a workplace, said:

> Now that he's obviously going to be 15 in October, I'm expecting him to be more independent and school are expecting him to be more independent and it's like thinking – for example, I keep thinking when [child's name] goes into Year 10, he's going to be doing work experience. So suddenly, […] it's obviously thinking, how's [child's name] going to get on when he has to go and work somewhere with somebody for a week? (Parent of Participant 6).

Similarly, one parent reported worrying about whether her son, who currently manages with ease to keep himself organised, will cope as the workload and expectations at school increase, "he seems to be quite good on that front, actually. I don't know whether it's because he's only Year 9 and maybe the big homework hasn't hit yet, because obviously, he still doesn't get loads of homework." Similarly, another parent reported, "I feel for me, it's one of the most difficult things in terms of how, and in terms of my kind of worries in the future because I don't think he copes very well with organisation." Parents were often aware that their children's planning and organisational abilities were not where they should be for their age but struggled to prioritise these skills in the midst of a period where there were already so many demands – both academic and social – on young people.

> You know, I probably need to start taking more risks and letting there be a few disasters […] stepping back, but then, at the moment it's still not working, but I suppose I've been more gentle than I should be at the moment, because he's right at the end of his A levels. He's got four more assignments, he's been under a lot of pressure, and it's not been the time, I don't think (parent of Participant 7).

The young person in question, also acknowledged that it was currently a time when the demands on him had been increasing, as he transitions away from the highly structured environment of school:

> So, what happened in different, like, secondary schools is we had, like, a sort of homework diary where we would keep everything in. And that's sort of fallen out of practice as you come into university, college or whatever, because people assume that you're just able to maintain it yourself, which for a large majority I am, I just write it down on a calendar. (Participant 7)

Parents and young people alike mentioned that multitasking was a particular challenge, which often was discussed in the context for the increasing need to engage in multiple activities simultaneously as parents and teachers start to gradually pull back supports:

> Like, doing my homework while, not doing my homework…like sorting out my bag while my Mum asks me to do something else as well and I said I can't do two things at once, but she said I need to learn to do multi-tasking (Participant 11).

Although the executive skills of the autistic young people involved were not considered to be universally behind those of non-autistic adolescents, described in one case as:

> He has actually always been, though, quite good on the organisation front, which is good. Maybe, actually, probably better than a lot of other boys his age, I think. He's quite on the ball with what day it is and what he needs and not losing things […] he's

> actually much better than his brother used to be at the same age (parent of Participant 6)

But, for the most part, the difficulties recounted were believed to be more than those simply of a disorganised teenager:

> I know that a normal child, […] an adolescent is unorganised, and whatever you have, but it's beyond that. It's not just an adolescent thing for [Child name], he's always been like that (parent of Participant 8).

Autistic identity and EF

When asked about whether they believed being autistic had anything to do with their ability to achieve the types of activities discussed, participants' beliefs varied a great deal. One young man pointed out that his ability to apportion cognitive resources was not on a par with his non-autistic friends, "My neurotypical friends are better at channelling and organising the energy that they put into each subject and try and equalise and balance that. Which I also do try and do" as well as acknowledging that being autistic meant his intense interests were sometimes detrimental to his everyday functional ability:

> I mean yeah, the things I've mentioned are symptoms of Asperger's syndrome. So, definitely. I suppose my level of attention can be quite biased towards things that I'm interested in … And yet there's key important aspects of life in which I can miss out on because I'm much more into focussing on these main interests. (Participant 4)

Similarly, another young man shared this view and highlighted that these difficulties might relate to the social difficulties they experienced:

> everybody is all so different […] there is still different neurotypicals, so it depends but my autistic traits do change these activities that we've been speaking about […] like planning and organisation and friendships, definitely, and friends and family, definitely (Participant 2).

Another, when talking about difficulties with multitasking claimed this was "because I'm autistic, mostly!" Another claimed being autistic meant, "I think I find it harder to be flexible in a change of plan." but when asked if it has any other consequences on his ability to plan or organise his life, he responded "not really". His mother however, felt differently:

> I think it's just having autism means it's a bit more difficult for him to organise himself or it's just more difficult to remember instructions. Like we were saying, a bit more difficult to multitask, I think. It definitely does affect organisation and time management

> and things like that. It definitely does. It's more apparent, I think, now that he's older, really. But he still does quite well, I think, considering (Parent of Participant 6).

Although, interestingly, she was unsure whether he had awareness of the multitude of ways in which autism likely affected his life:

> Now he's older he knows that he has autism, which is something he only became aware of, that he had a condition with a name, sort of thing, when he was in Year 7. But whether he knows exactly how it affects him or what areas of life it affects, I don't quite know about that (Parent of Participant 6).

Other participants were less keen to attribute any of their executive abilities to being autistic, with one young autistic man ensuring to emphasise that autism impacted on his social abilities but was *frustrated with everything in his life being attributed to being autistic*:

> Definitely not, no, but I don't think everything can be planted at the door of autism, because autism for me is more of a condition which affects social aspects of the thing. Like, for example, understanding body language, understanding facial expressions and everything else (Participant 7)

This sentiment was shared by another, who claimed the effect of autism was on how they were perceived by others, "My autism, I think it affects how much people like me." Another, when asked whether being autistic had anything to do with their executive abilities simply responded "no", when probed about whether the two things might be related for other autistic people, they again responded with a straightforward and emphatic "no". Another young autistic person was conflicted about how he perceived autism to relate to autistic people's abilities, simultaneously proud of the achievements of Aspies [people diagnosed with or who identify with the diagnosis of Asperger's Syndrome] and the difficulties that being on the autism spectrum present:

> So, I don't really see autism as a learning difficulty as such, I see it more of a like almost social impairment to, well to use a bad word as it were. But yeah, I don't really see Asperger's Syndrome that much as a learning difficulty, especially when you consider the sum of the works of Aspies when compared to neurotypicals is a lot wider. And so therefore I consider ... yeah Asperger's syndrome is definitely a learning difficulty. But it's a trade-off in the sort of knowledge that you're able to maintain, especially in terms of vocabulary and everything else. (Participant 7)

One parent intriguingly highlighted her *uncertainty about the direction of causality* between EF and autism:

> Whether it is [planning and organisational difficulties] that makes him autistic or being autistic that makes him unorganised, how could you define that? […] I think it makes, that [EF difficulties] makes him autistic rather the other way around. I think. I don't know. And this is my perception anyway (Parent of Participant 8).

Young autistic people in this study were often very accepting of *their cognitions and behaviours simply reflecting intrinsic aspects of who they are*, stating, "That's what I'm like", "I just think it's because I have a one-track mind […] that's nothing to be particularly ashamed about, it's just the way I work", or reporting that the behaviours they display are "just who I am, so I'll always be doing that". Other participants were less positive in their outlook about what being autistic, or more specifically the consequences of other people's perceptions and responses to their being autistic, meant to them:

> Well, I find difficult people saying I'm autistic. I dislike people saying, 'Hey you're autistic'. I find it very tagging, where you're tagging somebody for who they are or what they are, and I find it very, how do I say? Very insulting (Participant 9)

Although, references to the *individuality* of the way in which their own brain operates were common: "I just think it's because my mind, like subconsciously or consciously, just prioritises one piece of information over the other, and for that reason I'm not able to remember everything." Many of the autistic young people when asked about the interrelationship between EF and autism were keen to emphasise the individuality of experiences of autistic people and were often frustrated with the notion that because some autistic people struggled with executive tasks that all autistic people necessarily had these same struggles:

> I'm autistic and I can perfectly multi-task with books and stuff like that. But others, that's a different story because you can have a variety of autism, including those who hate multi-tasking and could probably throw a tantrum over that. […] but I think that's something that should be recalled because some autistic children and men and people, adults and children, can, they can, they can perfectly multi-task. Others, they can't. (Participant 9)

Another reinforced the importance of remembering that, despite sharing some commonalities with his autistic friends, they did not always have the same challenges, including those with executively-demanding tasks:

> The ironic thing about autism is all the friends I have that are autistic and me, we're all very, very different from each other, and sharing perhaps some symptoms that I used to characterise a diagnosis. But apart from that, very different (Participant 4).

A need for accommodations

A common thread running through many of the interviews concerned the accommodations and supports participants found to be helpful or those that they wished had been available. The most consistently referenced accommodation was that in many cases a parent, sibling or teacher was acting as an *external locus of EF* for the young people. One young person, who when asked about the support they received to overcome difficulties in school described how the school removes the need to deal with multitasking:

> We don't really multitask because most of the kids in the school I go to are autistic and could throw a fit at any moment, or anger at any moment when there's multitasking. So, it's always one task, and if you finish that task yes you go onto another task (Participant 4).

Similarly, at home, reducing the need for the young autistic people in this study to deploy and consequently to exercise their EF abilities was often described, one mother admitted:

> the truth is I do probably control his life. I'm trying to back off. I'm trying not to control it as much, but at the same time, I'm trying not to let him have disasters, so he loses confidence. As a mother, it's just a very, very narrow line to tread, you know (Parent of Participant 7).

Similarly, another conceded that she needed to provide constant prompts to her child to mitigate the fact that he would struggle to disengage from his intense interests: "he's already 14 years old; I still need to, in a kind of loose way, need to manage his time because otherwise he wouldn't do anything else but what he likes." Another young man admitted that he needed prompts from his parents to help him overcome his procrastination, "I just completely forget what I'm doing, or you know procrastinate on the phone, and on those instances, they do have to sort of prompt me again before I eventually do it." Another parent acknowledged the role of siblings in providing support, "[child's name]'s brother helps him structure his life and organise his life." Another young man spoke about his difficulty getting started on activities and the associated stress when external supports were not available:

> I might procrastinate a lot, I might say I will do something, and then I'll not do it later and have to do it all in 30 minutes, and have to do it all in one sitting, and then I will be very stressed out, then I'll be very stressed out, I am not very good at planning ahead of time (Participants 10).

There were many references to *helpful strategies that autistic people employed*, such as using tools to supplement their own cognitive abilities, for example, using "timers, lots and lots of timers", "she does use the phone for that", making and following lists ("Normally, I need a list and I keep it with me") or working hard to overcome some difficulties they have: "I try my best. I revise an awful lot, so I often write down things to try and remember them. And I'm actually doing okay, I mean I remember… yeah, no, I am actually doing okay". Notwithstanding the success of such strategies, contending with ineffective strategies that are suggested by others was sometimes an irritant: "I get frustrated when she comes up with ten or more completely different methods, all of which would seem exactly the same to me. And yeah, it just doesn't work really, planning, in general, for me."

On the topic of support provided in school, there was sense among participants that there was often a *mismatch between the accommodations made and the difficulties experienced.* One young person wished that more attention was focused on explicitly teaching organisational and time management skills: "Like, at what time should we revise, what time we shouldn't, what time we should revise this subject and that subject, that sort of thing." Another young person added that the benefits of teaching such skills would not be confined to those on the autism spectrum: "everyone should learn and be especially taught how to properly organise, because it could help." One parent expressed concern that if supports for executive difficulties were not provided within a window early in development, that the opportunity might pass by:

> However, I mean he still has his needs, so I'm kind of at the moment I am in the fight of, fight I will say, because I think it will take me that, to actually ensure that there is some help for him especially with this organisational skills because I think if we don't do it now it will become really ingrained and I'm very concerned (parent of Participant 8).

Similarly, some parents reported that often, the accommodation offered to their children with respect to exam settings and assignments was to allow them extra time, something that is not, according to these parents, helpful to somebody who struggles to manage deadlines:

> I think that all he really needed was somebody who could perhaps just have prompted him, "Don't sit there concentrating just on that, just move onto the next question, [child's name]." You know, not to give him any help with the work, but just to prompt him to be conscious of the fact that you've only got a certain amount of time, because that's difficult for him to really completely understand […] or a little alarm goes off or something for him as a reminder. (Parent of Participant 5)

This accorded well with another parent, who wished the supports that were available to her son focused on his difficulties, particularly focusing on interpreting instructions, knowing how to prioritise and reducing anxiety around each of these areas:

> I wish there was a lot more structured support in how to interpret questions and what you're being asked. How to prioritise what information […] He was allocated someone who was to buddy him for the first few weeks so that he knew how to get from classroom to classroom, but he rushes to be at the classroom for when the bell goes because he's so scared of being late or losing his way, and panics about every small, little thing (Parent of Participant 2).

As well as making aids available that act as an external source of EF, some parents reported that performance in school, which could be mistaken for weak EF, might sometimes instead be better explained by other difficulties that are not accommodated sufficiently well, such as motor coordination difficulties:

> I think he should have been allowed to use a laptop because he was diagnosed – they suspect he probably had dyslexia, but it was never conclusively diagnosed. But his writing was quite – if you'd look at it, you would think perhaps a 12-year-old had written it. So, that was something that, really, I think it was a shame that he wasn't allowed to use a laptop because I think he would have done much better had he been allowed to, but he wasn't (Parent of Participant 5).

Participants were not universally disappointed at the support provided at school, however, as some young autistic participants believed they received ample support: “I don’t think I can get any more help than I already have received. I mean I’m much better now than I used to be.” Another young man believed it would be more helpful if those supporting him gave him more space by “leaving me to do my stuff, […] so I don’t have people distracting me.”

Discussion

Autistic young people and their parents described the difficulties they experienced as they navigated the demands of their daily lives, difficulties that, in many cases, could be the behavioural consequences of EF difficulties or, in other cases, were not executive in nature but nevertheless play a role in the application of EF. The insights shared here include that EF has been operationalised by the availability of neuropsychological tasks some of the difficulties that autistic people describe are rarely measured. Another insight was that EF does not exist in a vacuum and, consequently, efforts to isolate EF ability from motivational, psychological or other cognitive abilities may be misguided. Furthermore, EF ability and the need to employ EF change over time but not necessarily in unison, the extent to which EF difficulties present challenges to an individual might therefore change at different stages in development. An individual's relationship with, and attitude toward, their cognitions and behaviours are complex and deeply personal, even if some of these cognitions bring challenges, researchers must then be cautious and sensitive in targeting cognitive differences with interventions. This is especially pertinent because the young people in the current study often discuss the mismatch between the support they are offered and the areas in which they believe themselves to require support.

The difficulties described by autistic young people and their parents included misunderstanding task instructions and, consequently, struggling to generate and execute a plan, which according to some participants, resulted in something of an inertia. Some young people explained that the inability to get started on achieving a goal resulted in increased anxiety and this rise in anxiety served to further compound the inertia they were experiencing. Some highlighted that external sources of EF, like prompts from parents, could offset the effects of anxiety-induced inertia and executive difficulties. Autistic inertia has been discussed by autistic advocates (M. Sparrow, 2016; Sullivan, 2002; van de Wettering, 2010) but has scarcely received attention in the empirical literature. Misunderstanding instructions has previously been discussed as a source of difficulty in everyday life (Preece & Jordan, 2010). The implication here is that poor performance on EF tasks might be because participants are acting on misunderstood instructions rather than having weakness in their EF. Conversely, time-limited tasks where an examiner is available to clarify confusion encountered with instructions are not suited to capturing the cycle of anxiety induced inertia described by participants.

Furthermore, the current study emphasised that conceptualising EF as a series of fixed abilities that can be meaningfully measured devoid of context, as is commonplace in experimental and psychometric paradigms, does not yield a complete picture of a person's functional ability. For instance, the extent to which an activity is motivating appears to have a sizeable impact on autistic adolescents ability and propensity to complete the activity, consistent with one previous report (Marks et al., 2000). This raises the possibility that the group differences often observed on executive tasks, which can be tedious and monotonous, might reflect, to an extent, group differences in participants' ability to overcome their lack of intrinsic motivation to fulfil the demands of an examiner. Which could be couched as a lack of social motivation among autistic participants (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012) or as an act of compliance from non-autistic participants (Kruglanski, 1975), depending on your viewpoint. Consequently, manipulating the degree to which a task is motivating to participants, rather than intentionally removing potentially motivating aspects of a task, might be more informative in understanding executive difficulties and their functional consequences.

Another prominent theme within the current study was that an individual's mental state (especially their anxiety) has a huge impact on their executive abilities, which accords with previous accounts from autistic people (Humphrey & Lewis, 2008; Ludlow et al., 2012) and experimental evidence from typical development (Horwitz & McCaffrey, 2008). Future research is therefore needed to better understand how autistic people's mental states affect performance on executive tasks, or how performing an executively-demanding task affects a person's mental state, to further elucidate the link between EF skills and functional ability. For instance, experimentally manipulating how anxiety provoking a task is or simply including a pre-post measurement of participants' mental state might help us to go beyond some of the inconsistencies within the EF literature at present.

The next main theme from this study was that participants and their parents perceived their EF ability as changing over time, developments that often brought about a greater desire for autonomy and independence, as is typical at the onset of adulthood. The increasing demands placed on individuals, at least in some cases, appear to outpace developments in EF, resulting in EF difficulties becoming more apparent as autistic people transition toward adulthood, this is a finding also reported by Cribb et al. (2018). The mismatch in increasing expectations for independence and perceived increases in EF ability is evidenced in other longitudinal studies that have shown that as autistic participants get older, their standardised scores on measures of adaptive behaviour markedly decline (Gotham, Pickles, & Lord, 2012; Pellicano et al., 2018; Pugliese et al., 2016; Salomone et al., 2018). Similarly, the gap between autistic and non-autistic participants' EF abilities have been shown to widen from childhood into adolescence in cross-sectional studies using teacher- (Kouklari, Tsermentseli, & Monks, 2018) and parent-reported EF (Rosenthal et al., 2013). This highlights autistic adolescents as a group who are under-represented in EF research.

There was a divide in the autistic participants' views was about how their EF difficulties were related to their identity, in general, or their identity as an autistic person, specifically. Some believed that the ways that they think and behave, including their executive idiosyncrasies, was an intrinsic part of who they were. Some young people easily arrived at this acceptance and others have worked hard to achieve it. Participants were not all equally as positive about this aspect of their entity and expressed a desire to be more like their non-autistic peers, consistent with another account (Humphrey & Lewis, 2008). The values autistic people place on their cognitive style has important ramifications for the development and implementation of interventions. Autistic people who value this aspect of their identity are unlikely to see the need for, or indeed are likely to be affronted by the idea of, an intervention that, from their perspective, has the eradication of a central tenet of their identity as its aim. Researchers who design interventions need therefore to work with young autistic people to ensure that such interventions align with their preferences and values and are not antagonistic to them.

Some young people were reticent to make any association between autism and EF, emphasising that the effects of being autistic, in their view, were primarily social in nature. There was variability between participants about their executive abilities and some participants were keen to point out this variability, stating that, in their experience, some autistic people had very good EF, but others struggled markedly in this domain. This is consistent with quantitative investigations of EF in autism, which have shown that indeed EF difficulties are not universal in autism (e.g., Pellicano, Maybery, Durkin, & Maley, 2006; Pellicano, 2010). Researchers must remain mindful that autistic people may perform well on executive assessments not because of a lack of sensitivity of the assessment but because they have strong executive skills. Likewise, EF interventions should be targeted at those who have demonstrated executive difficulties rather than at all autistic people.

At times, however, there were different perspectives shared between parents and their children about the overlap (or not) between EF and autism. This has important implications for research and practice, in particular, when one considers the scarcity of first-person accounts of EF difficulties in the autism literature and that the commonly employed parent-reported questionnaires about everyday EF difficulties are presented as de facto measures of EF difficulties. The methods of the current study were not suited to investigating whether the reports of young people themselves or that of their parents was more closely aligned with their performance on various EF assessments, or indeed, outside of the laboratory, but this is nevertheless an important avenue for future research.

Finally, participants and their parents shared the strategies and accommodations that they believed were effective in their lives. For instance, many parents and young people referred to their being many external sources of EF on which the young autistic person could rely, albeit sometimes to the irritation of the autistic person themselves. Both parents and young people conceded that it is often challenging for the siblings, parents and professionals who were acting as external sources of executive control to pull back this support because of the multiplicity of demands placed by adolescence. While the provision of support is valuable and valued, there is an associated risk. That is, if executive demands continue to be stripped from the environments of the autistic young people struggling with EF, they will not be provided with enough opportunities to exercise and develop these skills and they will continue to move along a developmental trajectory diverging ever further from their peers. The consequences of this were salient for the young people in this study, who like those in previous studies, worried about their about their ability to cope with the increasing demands that lie ahead for them (Browning et al., 2009; Cribb et al., 2018).

In conclusion, the insights of the autistic young people and their parents provide opportunities worthy of future research that could help redress the research and practice gap that currently exists about EF in autism. In future, autism researchers should; (i) develop assessments that tap the specific difficulties described by autistic people like getting started on a task or decoding multi-step instructions, (ii) measure performance on an assessment *and* measure the contextual factors like motivation and anxiety that are likely implicated in performance, (iii) acknowledge that some autistic people have very good EF ability and (iv) be sensitive toward those who do not want support or interventions that might change their cognitions, and in turn, change their identity. On a broader scale, the current study demonstrated the value of supplementing the positivist research into the cognitive processes of autistic people with phenomenological methods, not only to better understand the lived experience of those we study, but also to facilitate the development of an evidence base that more accurately captures these experiences and, as a result, better serves their needs.

Chapter 6

General Discussion

In this thesis, I sought to better understand the executive function (EF) abilities of autistic adolescents. Specifically, I aimed to test whether the conclusions drawn from traditional laboratory-based EF assessments have a bearing on the everyday lives of autistic young people. To address this aim, I examined whether the group differences between autistic and non-autistic participants on EF assessments commonly reported upon in the literature result from differences in the latent executive constructs underlying task performance or task-specific, non-executive characteristics (Chapter 2). I also assessed whether the group differences in EF ability often reported emerged on a representatively-designed measure of EF (Chapter 3). I further investigated how well individual differences in performance on both a systematically-designed and on a representatively-designed EF task would generalise to other metrics of real-world functional adjustment (Chapter 4). Finally, I sought to understand the personal and parental perspectives about the executive abilities of autistic young people and the realities of these (dis)abilities on everyday functioning (Chapter 5).

In this Chapter, I begin by summarising the main findings from the empirical studies presented in this thesis. Second, I discuss the contributions these studies make toward our understanding of EF in autism. Then, I describe the limitations inherent in the research presented herein. Finally, I outline some future directions for this programme of research that could both improve the generalisability of research findings and better elucidate the mechanisms underlying atypical performance on EF assessments.

Summary of the main findings

Systematically-designed tasks. Our ability to decipher the myriad neuropsychological studies of EF in autistic people has been hampered by the task impurity problem (Snyder et al., 2015), the sheer diversity of tasks and small, inconsistently matched samples. In Chapter 2, I therefore sought to compare the EF ability of autistic and non-autistic young people on latent EF variables to circumvent the task impurity problem. I used a battery of tasks that is standardised and manualised, ensuring that like-for-like replications are feasible, and that collected data and materials were open source to facilitate cumulative improvements in statistical power. The analysis presented in Chapter 2 demonstrated that autistic adolescents had significant difficulties on the global executive composite of moderate effect, increasing our confidence that the often-reported EF difficulties among autistic people are not an artefact of individual EF assessment paradigms but reflect difficulties in underlying EF constructs.

This conclusion was complicated somewhat by two patterns of results in Chapter 2, where I found no significant differences on any of the three factor scores representing subcomponents of EF (fluency, cognitive control, working memory) in addition to discordance between the pattern of group differences on individual EF assessments and the latent variables representing the underlying EF constructs. I suggested four possible explanations for these discordant findings. First, it is possible that the tasks included in each domain are not reliably tapping the same construct, but that shared variance represents the non-executive commonalities of completing laboratory assessments, at least among autistic participants. Second, perhaps instead the relative executive difficulties that autistic people readily display in childhood and adulthood diminish or disappear in adolescence because of a drop in executive abilities among autistic *and* non-autistic adolescents owing to the structural re-organisation of neural architecture that accompanies the onset of puberty (Demetriou et al., 2018). A drop in ability among both samples could render these particular EF assessments insensitive to difficulties at this developmental stage. A third possibility is hinted at from the pattern of tasks where autistic participants under-perform; tasks where the responses are more open-ended, and less structure is provided to the participant. Under-performance on these tasks aligns with previous accounts that autistic people perform especially poorly in situations that are ill-structured (S. J. White, 2013). Finally, it remains possible that group differences exist on at least some of the factor scores from the assessment battery but these group differences were of smaller magnitude than the effect sizes this study was powered to detect. One implication is that larger samples were needed to more definitively assess whether these differences could reliably be found. To that end, the data were deposited in a public repository to allow future investigations to build on these data.

Representatively-designed tasks. Further implications arose from Chapter 2, including that many neuropsychological EF tasks may become less sensitive during adolescence, they might not assess many components of EF that autistic people struggle with and, consequently, may not be that informative about their everyday functioning. Chapter 3 therefore sought to test whether the significant group differences in EF demonstrated in Chapter 2, and in the extant literature, were apparent on a representatively-designed task, adapted from Schneider et al. (2016), which more closely resembled an executively-demanding task of everyday life. Autistic participants underperformed, relative to their age- and ability-matched non-autistic counterparts, on three of the four dependent variables derived to assess performance and, on a composite variable. A more stringent analysis, correcting for multiple comparisons, continued to show that autistic participants underachieved on this task on two of the metrics assessed.

The theory that autistic people struggle particularly on tasks that are ill-structured (S. J. White et al., 2009) offers a testable hypothesis; that autistic participants should demonstrate greater difficulties on this everyday task than on the constrained tasks within a neuropsychological battery. The difficulties measured by the composite score on the representatively-design task (*d* = .54) were comparable with the difficulties measured by the composite score on the systematically-designed task (*d* = .51). This hypothesis was therefore not wholly supported. Nevertheless, demonstrating similar effect sizes in both assessments bolstered our confidence in the conclusion that the executive difficulties reported in the literature previously, appear to manifest in more true-to-life settings for young autistic people.

There were, however, some findings within Chapter 3 that complicated this interpretation. First, autistic participants did not commit more of the types of errors that were pre-specified as being a potential indicator of executive difficulties (such as re-boiling the kettle multiple times) than their non-autistic peers on this task, contrary to my hypothesis. It is difficult to reconcile this null finding with the lower scores on the composite variable and greater levels of difficulties on traditional EF assessments but a second finding from the Chapter might hold a possible clue. That is, autistic participants rated themselves as being less motivated by and less likely to enjoy engaging in this task. One conceivable explanation is therefore that the findings that autistic participants completed fewer task goals, completed goals to a lesser quality and executed less-involved plans were, at least in part, due to a difference in the thoughts and feelings they had about completing the task. Autistic and non-autistic participants were, however, no different in their ability to appraise their own performance, undermining previous suggestions that autistic people may lack insight into their own cognitions (U. Frith & Happé, 1999; Zahavi, 2010).

The generalisability of executive difficulties. As outlined in Chapter 1, demonstrating significant group differences on representatively-designed assessments is not a sufficient criterion for determining the ecological validity of an assessment. This is because increasing task representativeness may unintentionally reduce the internal validity of that assessment by introducing non-executive demands (e.g., sensory, social, motor) or by introducing measurement error (e.g., inaccuracies in the coding scheme or unanticipated events such as a jammed stapler) that drive variability in task performance and, in extreme cases, may stifle the EF-associated variability altogether. In such cases, individual differences in performance on a representatively-designed task might not relate to individual differences in anything of significance in the real world and so might lack ecological validity or, at least, be less ecologically valid than a systematically-designed task. Chapter 4 therefore assessed the generalisability of performance on the systematically- and on the representatively-designed EF assessments described previously to other metrics of everyday adjustment or, in other words, assessed the veridicality of these assessments.

To assess the generalisability of these EF assessments, I tested whether individual differences on a composite score for each measure were related to individual differences on four self-reported questionnaires that probed aspects of participants' everyday adjustment, including their quality of life, anxiety, everyday executive difficulties and self-perceived disability. When the relationship between the composite scores from the representative and systematically-designed EF tasks and each of these four questionnaires were assessed separately, the outcomes differed slightly, as follows. Performance on the systematically-designed assessment along with individual differences in general cognitive ability were significantly related to individual differences in everyday executive difficulties but not any of the other three measures of everyday adjustment. Individual differences on the representatively-designed EF assessment were related to variation in anxiety and quality of life, and although variance on this task alone did not significantly explain unique variance in everyday executive abilities, when it was enterred into a regression model combined with variance in participant age it explained significant variance in everyday executive abilities. In both cases, this was true only for autistic participants.

To test the relative generalisability of these assessments, individual differences on both assessments were entered within the same stepwise regression models. Significant models were found that predicted individual differences in everyday executive difficulties, anxiety and quality of life. A model including the systematically-designed task alone best predicted individual differences in everyday executive abilities. A model including the representatively-designed task alone best predicted individual differences in anxiety. Finally, a model including both assessments best predicted individual differences in quality of life, explaining an additional 3% of variance over and above a model with the representatively-designed task only. The implication of these results is that systematically- and representatively-designed assessments demonstrate veridicality depending on the real-world metric they are assessed against; difficulties on neuropsychological tasks do appear to map onto real-world occurrences of executive difficulties whereas an everyday executive task is informative about everyday executive abilities, anxiety and quality of life. The proportion of variance in the self-reported questionnaires about participants' everyday lives, explained by either the systematically- or the representatively-designed assessments was, nevertheless, reasonably small. One possibility for this, was that neither task was optimised to detect the specific types of executive difficulties autistic people report themselves as having.

Personal and parental perspectives on executive abilities. Chapters 2, 3 and 4 demonstrated that autistic people have executive difficulties, with effect sizes ranging from small to medium, and that these difficulties affect their everyday lives. These effect sizes neither map well onto the profound executive difficulties autistic people often report (Cribb et al., 2018; Davids et al., 2016; DePape & Lindsay, 2016) nor do they compare with the comparatively larger effect sizes regularly reported in questionnaire studies (e.g., Kenworthy et al., 2005; R. A. Lawson et al., 2015). One plausible explanation for the mismatch between questionnaire- and performance-based measures, is that the EF tasks used in autism research, which were originally intended to assess patients with frontal lobe lesions, may lack sensitivity to the subtler and sometimes qualitatively different difficulties experienced by autistic people, which may be better served by questionnaires assessing everyday EF. Furthermore, the autistic young people in Chapter 4 demonstrated that they did not lack insight into their difficulties, which poses a question: why has their voice been hitherto absent from the literature characterising EF difficulties in autistic people?

In Chapter 5, I therefore directly asked autistic young people and their parents about their executive abilities and the functional consequences of the difficulties they experienced in their day-to-day lives. The qualitative insights within this chapter do not straightforwardly map onto our theoretical models of EF and pinpoint how the operationalisation of EF, through our selection of assessments, have left many of the difficulties described by autistic people largely unmeasured in previous investigations.

For instance, autistic people and their parents highlighted that the ability to deploy EF is context dependent for autistic people, as it is for non-autistic people. This deserves an especially sharp focus in assessing the EF of autistic people because contextual factors, such as an individual's level of anxiety, motivation or sensory overload, might be drastically more changeable across situations and consequently have a sizeable impact on the variability on EF task performance for autistic people. Furthermore, this Chapter reported that difficulties understanding task instructions often led participants to feel uncertain about whether they were completing a goal as was intended, and this uncertainty, according to participants, caused heightened anxiety and sometimes induced an inertia. Standard measures of EF typically do not capture participants' understanding of instructions, their affect during the activity or their disposition toward completing the task. Finally, autistic young people were mixed in their views on whether their being autistic and having EF difficulties were related, highlighting the need for researchers to tread carefully and to develop interventions that attempt only to target attributes that autistic people believe to be well-placed targets.

Limitations

In this thesis, autistic participants' executive abilities were compared with those of age- and ability-matched non-autistic participants, in case-control studies (Chapters 3 and 4) and in individual differences analyses (Chapter 5). It is therefore beyond the scope of the analyses presented here to determine whether tasks with greater ecological validity differ with respect to their ability to differentiate between an autism-specific profile of EF difficulties and profiles of EF difficulties observed in other developmental or psychiatric conditions, as has been suggested previously (Kenworthy et al., 2008). That is, this thesis does not speak to the discriminant validity issue in EF assessment, introduced in Chapter 1.

On a related note, the participants in the studies presented within this thesis were cognitively able (i.e., had IQ scores above 70). This was to ensure participants understood the task instructions, that completing the breadth of testing undertaken was feasible and because I was specifically interested in the EF abilities of autistic people without intellectual disabilities (ID). When cohorts of autistic people who span the entire IQ range are considered, general intellectual ability is a powerful predictor of outcomes (Bal, Kim, Cheong, & Lord, 2015; Gillberg & Steffenburg, 1987; Howlin, Moss, Savage, & Rutter, 2013) but the outcomes of those without additional ID are highly variable and difficult to predict (Anderson, Liang, & Lord, 2014; Howlin et al., 2004). EF is one candidate that might be related to these variable outcomes and it is for this reason that I sought to better understand the EF of this specific group. One limitation of my research is nevertheless that we cannot generalise the findings presented herein to young autistic people who span the entire range of intellectual abilities.

Furthermore, there were not a comparable number of males and females in the studies described in this thesis because of the difficulty recruiting cognitively able autistic females. The sex ratios presented in the quantitative studies in this thesis (Chapter 1 = 3.1:1, Chapter 2 = 2.3:1, Chapter 3 = 3:1) are representative of recent estimates of autism diagnosis (Loomes et al., 2017) and are an improvement on many previous investigations. We must nevertheless be cautious in generalising the conclusions of these studies to autistic women and girls, who were under-represented in this research.

Despite the widely-held (Szameitat, Hamaida, Tulley, Saylik, & Otermans, 2015), but not empirically demonstrated (Strayer, Medeiros-Ward, & Watson, 2013), position that females have superior multitasking abilities to males there are no solid theoretical grounds to investigate sex-based differences in executive abilities. Investigating sex-based differences in the absence of valid theoretical grounds only serves to reinforce and legitimise stereotypical gender norms that are not scientifically justified, so-called 'neurosexism' (Fine, 2014). That is not to say, that gender-based differences could not exist in the abilities studied in this thesis, but should they exist, these differences likely materialise through cultural means. Previous empirical investigations of this nature were inconclusive (Kiep & Spek, 2017; M.-C. Lai et al., 2011). Regardless, the studies presented in this thesis were not designed to assess this possibility.

The autistic and non-autistic samples within this thesis were tightly matched groupwise for socioeconomic status, age, gender distribution, verbal IQ, non-verbal IQ and full-scale IQ. Yet this does not entirely preclude the possibility that individual differences in these abilities related to the group effects demonstrated within this thesis. There are two possible issues in drawing this conclusion. First, autistic people have been shown to have uneven IQ profiles comprised of relative peaks and troughs in performance (e.g., Joseph et al., 2002) and by matching groups based on total and subtotal scores there is a danger the groups will not be well-matched on any of the individual abilities assessed (Jarrold & Brock, 2004). This is offset somewhat in the studies presented within this thesis because groups were matched on verbal, nonverbal *and* full-scale IQ. Second, group matching does not completely control for the possibility that within-group variability on these factors is related to between-group differences on the outcomes we are interested in, in this case, executive ability (Neuhaus & Kalbfleisch, 1998). One approach to deal with this is to also co-vary the matched-for variables in the statistical tests conducted. Unfortunately, I had insufficient sample sizes to achieve adequate statistical power to achieve this in these studies, and so we must be cautious in our interpretation that group differences are a consequence of diagnostic group alone.

Many of the parents of the autistic young people involved in this research reported that their children had a range of co-occurring conditions and many were taking psychoactive medication at the time of testing. This represents a challenge in our ability to infer the extent to which participants' performance was related to autism, to a co-occurring condition, or to medication they were taking at the time of testing. Participants with co-occurring conditions were not excluded from analyses in this thesis because of the difficulties inherent in diagnosing co-occurring conditions in autism and inconsistencies in diagnostic practices (e.g., dual autism and ADHD diagnoses were officially recognised only in the most recent version of the diagnostic criteria).

One limitation of the null hypothesis significant testing framework, used here to determine the evidence for executive difficulties among autistic relative to non-autistic adolescents, is that when we fail to reject the null hypothesis, we cannot assume that the null hypothesis is true, or in this case, that autistic participants do not have EF difficulties. This is especially true because, although the sample sizes in the quantitative analyses presented in this thesis (ranging from 31 – 42 participants) were often larger than many previous investigations, they were only large enough to detect effect sizes of at least moderate magnitude. Consequently, we must remember that the absence of group differences on metrics of performance on the executive tasks presented in this thesis does not rule out the possibility that the autistic participants did, on average, have more difficulties on these metrics, but their difficulties were of a smaller magnitude than was predicted and that these studies were powered to detect. Assessing larger samples, which have enough statistical power to detect small effect sizes, would improve our inferential ability in this respect. Furthermore, using Bayesian statistics would allow us to infer whether the lack of group effects presented, represent intact executive ability or instead whether they represent insufficient evidence to support such an inference.

While performance on the assessments of EF investigated in this thesis were shown to be related to individual differences on other measures of everyday adjustment, one weakness of this study is that the veridicality of these assessments was only assessed against a limited number of self-reported, real-world outcomes, which limits our ability to infer the veridicality of these assessments in three important ways. First, it might be that had other self-reported metrics of real-world adjustment been employed, the demonstrated veridicality may have differed. Second, third party informant-reported measures may have demonstrated a different pattern of veridicality. Third, self- and informant-reported questionnaires are subject to bias and so we must be mindful of this caveat when interpreting the ecological validity of the measures presented in this thesis.

Besides, while the representatively-designed task within this thesis had more representativeness with respect to one everyday executively-demanding task, namely preparing food, drinks and materials for an event, than traditional neuropsychological assessments, it must be noted that this assessment may not represent other real-world executive tasks well. Furthermore, the assessment used was not truly representative of life outside of the psychology laboratory, for example, an examiner was present, and video recorded the task, it was tightly controlled across participants, the event was imagined and not a real study group. Each of these may limit the extent to which performance on this specific task does represent daily executively-demanding tasks.

Throughout this thesis, I have referred to the developmental course of executive abilities. For instance, I examined the relationship between age and performance on a systematically-designed (Chapter 2) and on a representatively-designed (Chapter 3) EF assessments in cross-sectional samples. I spoke with autistic young people and their parents about their perceptions of their changing EF abilities (Chapter 5). A more robust approach both to assess the role development plays in changing EF abilities would be to assess the same participants multiple times through development, in a longitudinal design. This was not feasible within this thesis but will be an important approach for future work in this area.

Finally, the coding framework applied to the representatively-designed task had a number of limitations. First, the coding framework did not account very well for differences in the efficiency with which participants completed the task. For example, a participant who completed all steps of the task by walking to and from the rubbish bin with each individual tea bag did not score any differently to a participant who completed each step but who moved the rubbish bin to be beside the cups of tea, saving time and demonstrating superior planning skills. Measuring efficiency, as well as accuracy, within a task that is, by definition, unconstrained, poses significant methodological challenges. This is because, unlike many traditional EF tasks, there are multiple possible routes to correct performance in naturalistic tasks, especially when a participant is required to apportion their time and chain multiple demands together. Second, groups may have differed on the frequency of dysexecutive errors that were subtler or qualitatively different than those captured within the list of possible dysexecutive errors in the coding frame applied. Third, it was challenging to categorically code a given behaviour as optimal or suboptimal because, in many instances, the extent to which a behaviour was optimal or not was relative depending on what a participant elected to do during another aspect of the task. Fourth, it was not possible to decipher if any errors committed by participants were the result of misunderstanding the task instructions rather than weaknesses of EF. Future studies should therefore compliment manual coding of videos with tracking of participant's movements within the task and should seek participants' appraisal of their performance to contextualise whether a given behaviour is best considered as inefficient, as an executive error or as a misunderstood task instruction.

Implications and future directions

As reported in Chapter 1, the literature pertaining to EF in autism is extensive. This area of research had its roots in attempts to determine whether a dysexecutive syndrome had developmental primacy over autistic traits. The idea that executive dysfunction underpins the development of autistic traits is no longer widely endorsed. However, research focusing on the EF of autistic people continues nevertheless, based on the blueprint set out by these early investigations, using neuropsychological tasks to test whether autistic samples underperform, on average, relative to non-autistic samples. I have argued, throughout this thesis, that this avenue of research now has limited utility and for EF autism research to be impactful, we must undergo a sea change and move beyond research designed to detect average group differences on abstract tasks with uncertain validity.

Researchers must now concentrate our collective effort on understanding how individual differences in EF ability impact on the life chances of autistic people, determine the factors associated with EF difficulties, and provide a rigorous evidence base for strategies that both improve EF skills and offset the consequences of non-executive factors that precipitate these difficulties. To do that, we need assessments that; (i) have prognostic power with respect to real life outcomes, (ii) are feasible to administer in the laboratory or clinic, (iii) are sensitive enough to identify which autistic people stand to gain from intervention and (iv) can be a primary outcome variable in randomised controlled trials of interventions. On that basis, I outline five recommendations for EF autism research below.

1. Validate systematically-designed assessments against real-world behaviour. In this thesis, I have demonstrated despite not representing the demands of daily tasks especially well, relatively pure assessments of specific cognitive abilities, like those included in the NIH-EXAMINER battery, hold some explanatory power with respect to autistic people's self-reported everyday executive difficulties. This finding justifies the continued use of this specific battery and suggests that difficulties on similar neuropsychological tasks are likely related to experiencing executive difficulties outside of the clinic or laboratory. Although, each assessment should be validated against the real-world behaviour over which one hopes to generalise its conclusions. The justification for the effort involved in validating tasks against various metrics of real-world function is justified given that validated measures are then adopted and used widely. The advantage to this practice comes from the relative ease with which neuropsychological tasks can be administered, relative to representatively-designed tasks which mean a validated neuropsychological task is comparably more feasible to administer in the laboratory or clinic. Importantly, the ecological validity of EF assessments can be population specific (Evans et al., 1997; Kenworthy et al., 2008). Measures should therefore be validated against real world outcomes of interest for autistic populations before being adopted for this means.

The NIH-EXAMINER was originally devised to assess EF in neurological patients and so it remains possible that it lacks sensitivity to the individual differences that exist within a typically developing sample and this lack of sensitivity is perhaps why it lacks ecological validity in this sample. Also, robust cognitive paradigms such as those included in this battery become popular because of their ability to establish reliable experimental effects, one reason for the reliability of these experimental effects is low between-participant variability, a quality that renders them less than ideal in correlational analyses (Hedge, Powell, & Sumner, 2017). Avenues for future investigation therefore include assessing the generalisability of performance on executive tasks or batteries of tasks that have greater sensitivity to individual differences across the entire range of ability and have greater between-participant variability.

The EF assessments in this thesis were compared to self-reported questionnaires in part, because a goal of my research was to understand how performance related to individual's own appraisal of their everyday functioning, in part, because they were feasible within the constraints of PhD research. The measure of self-perceived disability used was designed to be employed across a range of psychiatric and developmental conditions and apparently lacked sensitivity to the precise difficulties autistic people describe. Future studies should assess the validity of EF tasks against measures that are sensitive to the functional difficulties encountered by autistic people, for example, the Vineland Adaptive Behaviour Scales (S. S. Sparrow et al., 2005) or the ICF core sets (Bölte et al., 2018).

Questionnaires are subject to a catalogue of biases (Choi & Pak, 2005) and so a further worthwhile avenue for validating EF assessments would be to determine whether performance on a given task is related to participants' daily behaviour. Emerging methodologies offer opportunities to facilitate this. One methodology, known as, experience sampling or ecological momentary assessment, allows researchers to send prompts to participants' phones at regular intervals to report on their behaviour at the very moment they receive a prompt (Stone & Shiffman, 1994). The advantage to this method is that participants report on snapshots of their behaviour as they engage in their daily routines, which can be built up over a series of time points to better capture the variability of their behaviour, as opposed to questionnaires, which elicit subjective median estimates of previously engaged-in behaviour. This methodology also elicits reports that are not subject to participant distortions because of memory failures or expectancy bias. Illustrative examples of the utility of this methodology in validating laboratory assessments include demonstrations that laboratory measures of mind wandering (McVay, Kane, & Kwapil, 2009) and working memory capacity (Kane, Brown, et al., 2007) were related to participants' reports of mind-wandering in daily life.

Another approach that has potential in validating laboratory assessments against real-world behaviour is conducting structured observations of participants in their daily environment for a fixed period of time. For example, conducting structured classroom observations based on a coding frame to record instances that demonstrate difficulties in, for example, self-control or inhibition (Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; Whitebread et al., 2009).

2. Create assessments that represent the conditions over which you wish to generalise. The representatively-designed task described within this thesis demonstrated significant group differences and superior generalisability to the systematically-designed task, at least with respect to the measures of real-world functioning used in this research. On that basis, representatively-designed tasks seem best placed to fit the needs of research that is measuring EF with the aim of making claims about real-world functional adjustment, according to the research presented here. Recall that representativeness is not a term to describe whether a task contains 'real world' conditions, which is a poorly specified construct, but that it involves, first, a specification of the specific conditions over which one hopes to make a generalisation and, second, a specification of how those conditions are included within a task (Brunswik, 1948; Hammond, 1998). The task described in this thesis represented some conditions over which one might wish to make generalisations, such as, apportioning time; generating, monitoring and executing a strategy; chaining multiple goals together; and executing a plan in the context of sensory or anxiety-provoking conditions. These conditions might not all be represented equally well by this task and it certainly does not represent every condition one might wish to generalise to. Researchers aiming to understand how differences in EF abilities make the lives of autistic people more or less difficult therefore must first specify the conditions they hope to generalise the results of an assessment to and then to specify how a given task fits these specifications. Then they must also validate the task by demonstrating that performance on the task generalises to other metrics of real-world functioning.

Assessing participants on naturalistic tasks is challenging. Codifying the messy, relativistic and sometimes ambiguous behaviour that people produce in the service of real-world goals is not straightforward. Manually setting up a testing space to be identical for each participant, particularly when the task stimuli are perishable, as was the case in this study, is time consuming. Replenishing the materials used to complete the task can be expensive. Coding participant performance manually is time consuming and somewhat subjective. There are many instances in research and, especially in clinical practice, where tasks such as the one employed here would not be feasible. There are a number of strategies that may be useful in increasing the feasibility of conduct tasks with representativeness.

First, virtual reality also offers a real opportunity for representatively designed tasks, depending on the conditions one is seeking to represent. For instance, if the conditions you wish to represent in a task relate, for example, to navigating three-dimensional space, self-guided goal setting and chaining multiple goals together then assessing these in virtual reality can remove many of the barriers to representatively-designed tasks mentioned above. In fact, a number of such tasks have been developed to assess EF in various neurological patient populations, such as, for example, in a virtual library (Renison, Ponsford, Testa, Richardson, & Brownfield, 2012), virtual supermarket (Grewe et al., 2014; Klinger et al., 2006), virtual apartment (Zalla, Plassiart, Pillon, Grafman, & Sirigu, 2001) and in a virtual office (Jansari et al., 2014). Similarly, gamification represents a significant opportunity to efficiently collect large amounts of data, especially if this can be conducted online (Hyde et al., 2016). While appealing, these types of assessment will not however represent situations in which moving in veridical space, executing a plan in the context of sensory stressors or when simultaneously managing social demands is required. In situations where the aim of an assessment is to represent these attributes then there is no substitute for assessing a participant in contexts that closely mirror daily life.

Second, technology offers opportunities to automate the coding of participant behaviour in situations representative of everyday life which could serve to drastically reduce the investment of time and subjectivity currently involved in assessments of this nature. For example, advances in eye-tracking technology mean that it is now feasible to measure participants' eye gaze as they engage with the world around them (Ye et al., 2012). Developing paradigms that capture and analyse the rich data generated from eye gaze during everyday tasks could bridge the gap between the convenience of virtual tasks and the veridicality of the real world. An example, employing a similar technique, involved tracking participants' eye gaze as they made sandwiches in the context of many distractor items, where anticipatory saccades toward relevant items were taken as proxies of task-related planning (Hayhoe & Ballard, 2005; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003).

Third, movement capture, GPS and indoor positioning have becoming increasing accurate and freely available (Arsan & Kepez, 2017). These provide opportunities to assess the efficiency and accuracy with which a participant engages in executively demanding tasks in everyday life. For instance, allowing researchers to measure navigational, wayfinding or strategic planning abilities in veridical space rather than with virtual or paper-and-pencil analogues. Alternatively, video annotation software whereby researchers can make detailed annotations that capture the amount of times participants, for example, switch between discrete subtasks, return to previous subtasks, re-read task instructions or spend planning before making their first action could yield useful information about differences in the application of executive skills to representatively-designed tasks (Kipp, 2012; T. Schmidt et al., 2008). While this methodology does not overcome the need to invest time to capture participant performance it does however have the potential to provide information about the efficiency, as well as the accuracy, of executive processes as there are applied online in real environments.

3. Triangulate objective assessment and subjective reports. Within this thesis we saw that the qualitative insights from autistic young people and their parents do not straightforwardly map on to our theoretical conceptualisations of EF and draws the way we operationalise EF, through the selection of available assessments, squarely into question. If the types of executive difficulties described by autistic people, such as self-guided multitasking, initiating execution of a plan while uncertain about the instructions or while decoding the social nuances of the situation, remain absent from the assessments we typically employ, then we will continue to underestimate the extent and consequences of executive difficulties among autistic people.

An important implication of this finding is that we must undertake a thoroughgoing rethink of standard psychological paradigms which focus on the speed or accuracy with which a participant responds to a given stimulus to assess cognitive processes like EF (Jack & Roepstorff, 2002). According to Jack and Roepstorff (2002), "[subjective] experience is still regarded as a problem, rather than a resource ready to be tapped" (p. 334). Psychological scientists are often quick to highlight that participants can be, and often are, mistaken about their cognitive processes and the reasons for their actions (Nisbett & Wilson, 1977), but according to Jack and Roepstorff (2002), understanding cognitive processes based solely on 'objective' observations of people's behaviour or images of their brain activity while they complete stimulus-response tasks is no less straightforward. To further our understanding, they argue, we need to engage in methodological triangulation in which objective measures of behaviour and neural activity are related to subjective introspective evidence.

Successful examples of this approach have involved assessing participants' ability to navigate a virtual environment while recording their functional brain activity during this task and supplementing this information with a participants' own verbal recall of what they remembered thinking during the task, recorded using a post-task, predetermined, step-by-step protocol (Spiers & Maguire, 2006). As well as providing the researcher with valuable, previously untapped insights into the cognitive mechanisms implicated in task performance, triangulating the objective performance and subjective experiences of task completion serve to democratise the study of EF difficulties in autism, such that researchers and autistic people pursue the research questions that matter most to autistic people. This may also reduce the paternalistic and, often medically and deficit-based, approaches to researching autism as a 'dysexecutive syndrome' and instead reframe the different patterns of ability in less value-laden terms. This may provide an opportunity to reconcile objective assessment with theoretical positions championed by autistic people whose lived experience of EF differences is described, not as a deficit, but as a Monotropism, where autistic people expend all of their attention at one focal point of interest and do not disperse their attention across numerous points of interest, in the same way as non-autistic people do (Murray, Lesser, & Lawson, 2005).

4. Elucidate the mechanisms the underlie everyday EF difficulties with experiments. The research presented within this thesis has gone some way toward uncovering that EF difficulties in autism are not merely artefacts of neuropsychological assessment but instead difficulties that manifest in tasks that mirror the demands of daily life, are related to other metrics of everyday adjustment and have real consequences, according to autistic young people and their parents. The research presented in this thesis did not, however, uncover the mechanisms through which these difficulties arise. For example, it is unclear whether the relative underperformance of autistic young people in the representatively-designed EF task is attributable to a reduced motivation to complete the task, to heightened anxiety in response to the task, to a difficulty decoding task instructions, to a difficulty generating an appropriate plan or to a reduced ability to execute a well formed plan, perhaps poor performance is instead the result of a specific difficulty simultaneously generating and executing a plan. Careful experimental manipulation can serve to uncover a mechanistic understanding of these difficulties.

A patent example from the literature comes from the study of patients with Action Disorganization Syndrome (ADS), a syndrome to describe patients who, following brain injuries that are more diffuse than those sustained by patients with Dysexecutive Syndrome, make cognitive errors on familiar tasks that require chaining multiple steps together to achieve a goal (Schwartz, Reed, Montgomery, Palmer, & Mayer, 1991). An elegant set of experimental investigations were conducted, in which participants with ADS were assessed under different experimental conditions, similar to the everyday activities assessed within this thesis, such as preparing a cup of tea, a sandwich, posting a card and wrapping a gift (Morady & Humphreys, 2011). The manipulations made by Morady and Humphreys (2011) were that participants had to (i) complete the activity themselves after hearing instructions, (ii) instruct the examiner to complete the activity, (iii) complete the activity themselves but with the aid of the task instructions, (iv) complete the activity but receiving feedback on each step as it was completed and (v) completing each of the component steps individually with individual instructions delivered for each step. This design allowed the authors to test whether difficulties completing the standard condition (i) were reduced in the condition where participants instructed the examiner to complete the task (ii) and so did not have to maintain the task schema while generating appropriate motor responses, did not have to engage in error monitoring (which is not necessary when the examiner executes the task correctly), or when competition for cognitive resources from antecedent or subsequent task steps is removed. The subsequent conditions (iii – iv) were included to assess separately the relative role of maintaining the task schema, engaging in error monitoring and competition from other task steps, respectively.

Autistic young people and their parents provided overwhelmingly consistent reports that executive abilities do not exist in a vacuum and cannot be universally deployed irrespective of context and an individual's mental state, drawing particularly on anxiety to demonstrate this. Conducting carefully-controlled experiments to assess the relative contribution of contextual and antecedent factors would further our understanding of EF in autism. An informative example from the literature involved experimentally manipulating the affective state of participants while they completed standard EF assessments (Horwitz & McCaffrey, 2008). In this example, a between-subject design was employed to assess university students on a measure of verbal fluency and the Trail Making Test either while being observed by a third party or on their own. Horwitz and McCaffrey (2008) demonstrated that the there was an effect of observer presence, which they refer to as a proxy for state anxiety, on overall performance and that there was a significant interaction effect between observer presence and participants' self-reported trait anxiety. A similar paradigm could be employed to assess whether trait and state anxiety have a comparable role in task performance in autistic and non-autistic participants or whether heightened state and/or trait anxiety or a qualitatively different relationship with task performance exists in autism.

As well as understanding the role of an individual's affective state, understanding differences in motivation toward tasks represents a good opportunity to disambiguate discrepant findings from the literature. When we, as researchers, request participants to complete a task, participants have thoughts and feelings about that request as well as the thoughts they have in the service of completing that task. At present, the tradition is to capture the latter form of cognitions only. Neuropsychological assessments have tended toward stripping content that might be motivating for some individuals, and not others, from tasks, to avoid confounding. This approach has advantages, but it also has pitfalls, participants still have thoughts and feelings about a task that involves responding to abstract stimuli, such as individual differences in competitiveness, reactivity to feedback or being differentially sensitive to experimenter bias and thus, the possibility that task performance is confounded with motivation is not wholly controlled. Second, participants might still differ with respect to how motivating the stimuli presented are, even in cases when abstract stimuli are used, for example, some participants are motivated by computers (which are used to present many common EF tasks) or by social interaction (which is almost always involved in data collection of this type). Asking a participant to rate their disposition toward completing a task, in the laboratory and in real world analogues, such as that presented in Chapter 3, and testing whether individual differences in performance differ by group, is one possible method to the bolster the inferences made from these tasks. Quasi-experimental design provides another, albeit more difficult to achieve, opportunity to disentangle the explanatory power of participant motivations and disposition towards individual differences in task performance. For example, grouping participants who are high and low on a construct of interest, such as competitiveness or conscientiousness, and assessing whether differences in these attributes are related to task performance or interact with diagnostic group. Experimental and quasi-experimental manipulations therefore provide a powerful framework to assess the mechanisms that underlie poor performance on everyday tasks and should be pursued in future research about everyday behavioural difficulties of autistic people.

5. Research EF within dimensional, rather than categorical, frameworks. The move away from attempts to investigate EF 'deficits' as a causal explanation of autism, or as an all-or-nothing ability that can be meaningfully measured without consideration for environmental or dispositional context, necessitates considering EF abilities within a dimensional framework, rather than the traditional categorical approach. For instance, the International Classification of Functioning, Disability and Health framework (ICF; WHO, 2001), which conceptualises disability as the complex interplay between an individual's impairment (e.g., EF difficulties), the activities in which they engage (e.g., the amount of executive abilities required by a given task), environmental factors (e.g., whether their difficulties are accommodated for, or the presence of stressors) and personal factors (e.g., motivation or personality) which, together, impact on the extent to which somebody participates in society around them (Rauch et al., 2012). We should therefore concentrate our efforts on understanding how each of the dimensions considered by the ICF are implicated in combining with executive difficulties among autistic people to more fully understand their functional (dis)abilities.

The National Institute of Mental Health (NIMH) recognised that the existing clinically-derived nosologies were impeding scientific progress and, to try to circumvent this problem, developed a new Research Domain Criteria (RDoC). The aim of RDoC was, for research at least, to supplant the diagnostic categories with a dimensional scaffold to better allow us to understand the connections between genes, neural circuitry and behaviour (observed and self-reported). Unlike DSM and ICD, it does not aim to detect relationships between each of these different levels of analysis and categories that have repeatedly failed to demonstrate reliability and validity. RDoC instead aims to build understanding of how the inter-relationship between genes, neural circuitry and behaviour explain the full continuum of experience on universal psychological domains. These domains are: (i) positive valence systems, (ii) negative valence systems, (iii) cognitive systems, (iv) systems for social processes and (v) arousal and modulatory systems (Cuthbert & Insel, 2013). Critically, RDoC hopes to enable understanding of the connections between these domains within the critical frame of neurodevelopmental and environmental influences (Insel et al., 2010).

By applying RDoC to our conceptualisation of autism, it becomes a brain-based condition resulting from a particular genetic make-up, which interacts with both developmental and, broadly defined, environmental influences to produce atypical brain circuitry. This atypical brain circuitry then creates patterns of variability on each of the five domains outlined above. By focusing on the interactions between brains, development, environment and each domain, we may increasingly understand the mechanisms through which a group of people, for example, have a particularly responsive valence system or why a certain cognitive profile emerges. The modification brought about by the framework, the theory goes, allow us to investigate why variability in these domains may be differentially disabling or enabling depending on the individual, environment and stage they are at in development. This is a significant deviation from the purely medical model conceptualisation that underpins DSM, which contends that autism is a neurological disorder resulting from a shared neuropathology underlying a shared cognitive profile which leads to an observable constellation of common behaviours (Insel et al., 2010). In theory, RDoC should allow us to target support toward the specific domains that require intervention and should not assume that any one intervention should predictably improve outcomes for all members of an ill-defined group.

The DSM diagnoses have historically classified only phenomena that produce impairment, whereas RDoC dimensions are conceptually agnostic in the phenomena they describe (Fein, 2016). Thus, autism, in this context, is multivalent – encompassing traits that are valued in some contexts (e.g. attention to detail) as well as those that are more universally devalued (e.g., aversive reactions to sensory information). This has important ramifications when conceiving of possible interventions through a biomedical lens, that is, ameliorating the impact of psychopathology may reduce not only universally devalued traits but it may also serve to reduce one's propensity for displaying valued parts of their identity (Fein, 2016). Approaching, EF difficulties within the RDoC therefore involves identifying those who struggle with EF, irrespective of their diagnosis, and understanding how these difficulties manifest and how they can be remediated. A worthwhile avenue for future investigation could therefore be to study the correlates and consequences of EF difficulties across the whole spectrum of (dis)ability and to focus our efforts on targeting support and intervention toward those who stand to benefit and who seek such supports rather than unilaterally applying an intervention based on a diagnostic category alone.

Concluding remarks

To conclude, the research presented within this thesis has increased our confidence that the EF group differences previously reported in the literature, with some degree of inconsistency, do relate to underlying executive skills and are not necessarily artefacts of individual tests. Furthermore, I have shown that autistic people struggle with an executive task more closely related to an everyday activity. Both the traditional neuropsychological and the increasingly representative task were related to autistic participants' everyday executive abilities, but the naturalistic task also relates to their anxiety and quality of life. Autistic young people and their parents provided insights into their executive difficulties that are rarely assessed in most investigations. I have made recommendations, on the basis of these findings, to further understand how individual differences in EF relate to the life chances of autistic people, to build a rigorous evidence base to ensure autistic people are afforded the best opportunities in life. Following these recommendations should ensure that research in this arena can increasingly answer the questions that are relevant to those for whom this research matters most – autistic people, their families and their allies.

References

Adamo, N., Huo, L., Adelsberg, S., Petkova, E., Castellanos, F. X., & Di Martino, A. (2014). Response time intra-subject variability: commonalities between children with autism spectrum disorders and children with adhd. *European Child and Adolescent Psychiatry*, *23*(2), 69–79.

Adams, N. C., & Jarrold, C. (2009). Inhibition and the validity of the stroop task for children with autism. *Journal of Autism and Developmental Disorders*, *39*(8), 1112–1121.

Adams, N. C., & Jarrold, C. (2012). Inhibition in autism: children with autism have difficulty inhibiting irrelevant distractors but not prepotent responses. *Journal of Autism and Developmental Disorders*, *42*(6), 1052–1063.

Agam, Y., Joseph, R. M., Barton, J. J. S., & Manoach, D. S. (2010). Reduced cognitive control of response inhibition by the anterior cingulate cortex in autism spectrum disorders. *NeuroImage*, *52*(1), 336–347.

Ahern, K. J. (1999). Pearls, pith, and provocation: ten tips for reflexive bracketing. *Qualitative Health Research*, *9*(3), 407–411.

Akbar, M., Loomis, R., & Paul, R. (2013). The interplay of language on executive functions in children with asd. *Research in Autism Spectrum Disorders*, *7*(3), 494–501.

Altgassen, M., Koban, N., & Kliegel, M. (2012). Do adults with autism spectrum disorders compensate in naturalistic prospective memory tasks? *Journal of Autism and Developmental Disorders*, *42*(10), 2141–2151.

Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357–376.

Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in Neurosciences*, *31*(3), 137–145.

Ambery, F., Russell, A. J., Perry, K., Morris, R., & Murphy, D. G. M. (2006). Neuropsychological functioning in adults with asperger syndrome. *Autism*, *10*(6), 551–564.

Ames, C. S., & Jarrold, C. (2007). The problem with using eye-gaze to infer desire: a deficit of cue inference in children with autism spectrum disorder? *Journal of Autism and Developmental Disorders*, *37*(9), 1761–1775.

Andersen, P. N., Skogli, E. W., Hovik, K. T., Geurts, H., Egeland, J., & Øie, M. (2015). Working memory arrest in children with high-functioning autism compared to children with attention-deficit/hyperactivity disorder: results from a 2-year longitudinal study. *Autism*, *19*(4), 443–450.

Anderson, D., Liang, J. W., & Lord, C. (2014). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *55*(5), 485–494.

Aoki, Y., Abe, O., Nippashi, Y., & Yamasue, H. (2013). Comparison of white matter integrity between autism spectrum disorder subjects and typically developing individuals: a meta-analysis of diffusion tensor imaging tractography studies. *Molecular Autism*, *4*(1), 25.

APA. (1980). *Diagnostic and statistical manual* (3rd ed.). Washington: American Psychiatric Association.

APA. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington: American Psychiatric Association.

APA. (2000). *Diagnostic and statistical manual, text revision* (4th ed.). Washington: American Psychiatric Association.

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

Arsan, T., & Kepez, O. (2017). Early steps in automated behavior mapping via indoor sensors. *Sensors (Switzerland)*, *17*(12).

Asperger, H. (1944). Die "autistischen psychopathen" im kindesalter. *Archiv Für Psychiatrie Und Nervenkrankheiten*, *117*(1), 76–136.

Auyeung, B., Taylor, K., Hackett, G., & Baron-Cohen, S. (2010). Foetal testosterone and autistic traits in 18 to 24-month-old children. *Molecular Autism*, *1*(1), 11.

Baddeley, A. (1992). Working memory. *Science*, *11*(5044), 556–559.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49A*(1), 5–28.

Baddeley, A., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation*, *8*, 47–89.

Baddeley, A., & Wilson, B. A. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, *7*(2), 212–230.

Bagatell, N. (2018). From cure to community: transforming notions of autism. *Ethos*, *38*(1), 33–55.

Bailey, A., Lecouteur, A., Gottesman, I., Bolton, P. F., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder - evidence from a british twin study. *Psychological Medicine*, *25*(1), 63–77.

Baines, A. M. D. (2012). Positioning, strategizing, and charming: how students with autism construct identities in relation to disability. *Disability and Society*, *27*(4), 547–561.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in south thames: the special needs and autism project (snap). *The Lancet*, *368*(9531), 210–215.

Bal, V. H., Kim, S.-H., Cheong, D., & Lord, C. (2015). Daily living skills in individuals with autism spectrum disorder from 2 to 21 years of age. *Autism*, *19*(7), 774–784.

Banich, M. T. (2009). Executive function: the search for an integrated account. *Current Directions in Psychological Science*, *18*(2), 89–94.

Barendse, E. M., Schreuder, L. J., Thoonen, G., Hendriks, M. P. H., Kessels, R. P. C., Backes, W. H., … Jansen, J. F. (2017). Working memory network alterations in high-functioning adolescents with an autism spectrum disorder. *Psychiatry and Clinical Neurosciences*, 73–83.

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.

Barkley, R. A., & Fischer, M. (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: self-reported executive function (ef) deficits versus ef tests. *Developmental Neuropsychology*, *36*(2), 137–161.

Barnard, L., Muldoon, K., Hasan, R., O'Brien, G., & Stewart, M. (2008). Profiling executive dysfunction in adults with autism and comorbid learning disability. *Autism*, *12*(2), 125–141.

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

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

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

Bastiaansen, D., Koot, H. M., Ferdinand, R. F., & Verhulst, F. C. (2004). Quality of life in children with psychiatric disorders: self-, parent, and clinician report. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(2), 221–230.

Bauman, M. L. (2010). Medical comorbidities in autism: challenges to diagnosis and treatment. *Neurotherapeutics*, *7*(3), 320–327.

Bayless, J. D., Varney, N. R., & Roberts, R. J. (1989). Tinker toy test performance and vocational outcome in patients with closed-head injuries. *Journal of Clinical and Experimental Neuropsychology*, *11*(6), 913–917.

Bellini, S. (2006). The development of social anxiety in adolescents with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *21*(3), 138–145.

Bennetto, L., Pennington, B. F., & Rogers, S. J. (1996). Intact and impaired memory functions in autism. *Child Development*, *67*(4), 1816–1835.

Berger, H. J. C., Aerts, F. H. T. M., Spaendonck, K. P. M. van, Cools, A. R., & Teunisse, J.-P. (2003). Central coherence and cognitive shifting in relation to social improvement in high-functioning young adults with autism. *Journal of Clinical and Experimental Neuropsychology*, *25*(4), 502–511.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336.

Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Research*, *1380*, 42–77.

Bettelheim, B. (1959). Joey: a "mechanical boy." *Scientific American*, *200*(3), 116–130.

Bhaskar, R. (1978). On the possibility of social scientific knowledge and the limits of naturalism. *Journal for the Theory of Social Behaviour*, *8*(1), 1–28.

Bíró, S., & Russell, J. (2001). The excution of arbitrary procedures by children with autism. *Development and Psychopathology*, *13*(1), 97–110.

Bishop, D. V. M., & Norbury, C. F. (2005). Executive functions in children with communication impairments, in relation to autistic symptomatology. i: generativity. *Autism*, *9*(1), 7–27.

Blair, C., & Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLoS ONE*, *9*(11), e112393.

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: a developmental psychobiological approach. *Annual Review of Psychology*, *66*(1), 711–731.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663.

Bleuler, E. (1911). Dementia praecox oder gruppe der schizophrenien. In G. Aschaffenburg (Ed.), *Handbuch der psychiatrie* (pp. 1–420). Leipzig: Franz Deuticke.

Bölte, S. (2014). The power of words: is qualitative research as important as quantitative research in the study of autism? *Autism*, *18*(2), 67–68.

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.

Bölte, S., Mahdi, S., de Vries, P. J., Granlund, M., Robison, J. E., Shulman, C., … Selb, M. (2018). The gestalt of functioning in autism spectrum disorder: results of the international conference to develop final consensus international classification of functioning, disability and health core sets. *Autism*, *online fir*, 136236131875552.

Bölte, S., Schipper, E., Robison, J. E., Wong, V. C. N., Selb, M., Singhal, N., … de Vries, P. J. (2014). Classification of functioning and impairment: the development of icf core sets for autism spectrum disorder. *Autism Research*, *7*(1), 167–172.

Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Journal of Experimental Child Psychology*, *33*(1), 87–110.

Boulter, C., Freeston, M., South, M., & Rodgers, J. (2014). Intolerance of uncertainty as a framework for understanding anxiety in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(6), 1391–1402.

Bramham, J., Ambery, F., Young, S., Morris, R., Russell, A. J., Xenitidis, K., … Murphy, D. G. M. (2009). Executive functioning differences between adults with attention deficit hyperactivity disorder and autistic spectrum disorder in initiation, planning and strategy formation. *Autism*, *13*(3), 245–264.

Brandimonte, M. A., Filippello, P., Coluccia, E., Altgassen, M., & Kliegel, M. (2011). To do or not to do? prospective memory versus response inhibition in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Memory*, *19*(1), 56–66.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Bristow, E. K. (2016). *Developing and piloting a new measure of executive functioning for children with autism spectrum disorder (ASD)*. UCL (University College London).

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of "hot" and "cool" executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337–349.

Browning, J., Osborne, L. A., & Reed, P. (2009). A qualitative comparison of perceived stress and coping in adolescents with and without autistic spectrum disorders as they approach leaving school. *British Journal of Special Education*, *36*(1), 36–43.

Brugha, T. S., McManus, S., Bankart, J., Scott, F., Purdon, S., Smith, J., … Meltzer, H. (2011). Epidemiology of autism spectrum disorders in adults in the community in england. *Archives of General Psychiatry*, *68*(5), 459.

Brunsdon, V. E., Colvert, E., Ames, C. S., Garnett, T., Gillan, N., Hallett, V., … Happé, F. G. E. (2015). Exploring the cognitive features in children with autism spectrum disorder, their co-twins, and typically developing children within a population-based sample. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(8), 893–902.

Brunswik, E. (1948). Systematic and representative design of psychological experiments: with results in physical and social perception. In J. Neyman (Ed.), *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* (pp. 143–202). Berkeley, CA: University of California Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.

Burack, J. A. (1994). Selective attention deficits in persons with autism: preliminary evidence of an inefficient attentional lens. *Journal of Abnormal Psychology*, *103*(3), 535–543.

Burgess, P. W. (1997). Theory and methodology in executive function research. *Methodology of Frontal and Executive Function*, (October), 81–116.

Burgess, P. W. (2000a). Real-world multitasking from a cognitive neuroscience perspective.

In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 465–472). Cambridge: MIT Press.

Burgess, P. W. (2000b). Strategy application disorder: the role of the frontal lobes in human multitasking. *Psychological Research*, *63*(3–4), 279–288.

Burgess, P. W., Alderman, N., Evans, J. J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*(6), 547–558.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M., Dawson, D. R., … Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, *12*(2), 194–209.

Cambridge Cognition. (2002). Cambridge neuropsychological testing automated battery (cantab). Cambridge Cognition.

Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is more: executive function and symbolic representation in preschool children. *Psychological Science*, *16*(8), 609–616.

Carmo, J. C., Duarte, E., Pinho, S., Marques, J. F., & Filipe, C. N. (2015). Verbal fluency as a function of time in autism spectrum disorder: an impairment of initiation processes? *Journal of Clinical and Experimental Neuropsychology*, *37*(7), 710–721.

Caron, C., & Rutter, M. (1991). Comorbidity in child psychopathology: concepts, issues and research strategies. *Journal of Child Psychology and Psychiatry*, *32*(7), 1063–80.

Carper, R. A., Moses, P., Tigue, Z. D., & Courchesne, E. (2002). Cerebral lobes in autism: early hyperplasia and abnormal age effects. *Neuroimage*, *16*(4), 1038–1051.

Carrington, S., & Graham, L. (2001). Perceptions of school by two teenage boys with asperger syndrome and their mothers: a qualitative study. *Autism*, *5*(1), 37–48.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404.

Cederlund, M., Hagberg, B. S., & Gillberg, C. (2010). Asperger syndrome in adolescent and young adult males. interview, self - and parent assessment of social, emotional, and cognitive problems. *Research in Developmental Disabilities*, *31*(2), 287–298.

Chan, A. S., Cheung, M. chun, Han, Y. M. Y., Sze, S. L., Leung, W. W., Man, H. S., & To, C. Y. (2009). Executive function deficits and neural discordance in children with autism spectrum disorders. *Clinical Neurophysiology*, *120*(6), 1107–1115.

Chan, R. C. K., Shum, D. H. K., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216.

Channon, S., Charman, T., Heap, J., Crawford, S., & Rios, P. (2001). Real-life-type problem-solving in asperger's syndrome. *Journal of Autism and Developmental Disorders*, *31*(5), 461–469.

Channon, S., & Crawford, S. (1999). Problem-solving in real-life-type situations: the effects of anterior and posterior lesions on performance. *Neuropsychologia*, *37*, 757–770.

Channon, S., Pratt, P., & Robertson, M. M. (2003). Executive function, memory, and learning in tourette's syndrome. *Neuropsychology*, *17*(2), 247.

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: data from the special needs and autism project (snap). *Psychological Medicine*, *41*(3), 619–627.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181–197.

Chen, E. Y. H., Shapleske, J., Luque, R., McKenna, P. J., Hodges, J. R., Calloway, S. P., … Berrios, G. E. (1995). The cambridge neurological inventory: a clinical instrument for assessment of soft neurological signs in psychiatric patients. *Psychiatry Research*, *56*(2), 183–204.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–238.

Choi, B. C. K., & Pak, A. W. P. (2005). A catalog of biases in questionnaires. *Preventing Chronic Disease*, *2*(1), A13.

Christ, S. E., Holt, D. D., White, D. A., & Green, L. (2007). Inhibitory control in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*(6), 1155–1165.

Christ, S. E., Kester, L. E., Bodner, K. E., & Miles, J. H. (2011). Evidence for selective inhibitory impairment in individuals with autism spectrum disorder. *Neuropsychology*, *25*(6), 690.

Clark, C., Prior, M., & Kinsella, G. (2002). The relationship between executive function abilities, adaptive behaviour, and academic achievement in children with externalising behaviour problems. *Journal of Child Psychology and Psychiatry*, *43*(6), 785–796.

Clark, L. A., Watson, D., & Reynold, S. (1995). Diagnosis and classification of psychopathology challenges to the current system and future directions. *Annual Review of Psychology*, *46*.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences 2nd edn*. Hillsdale, NJ: Erlbaum Associates.

Connor, M. (2000). Asperger syndrome (autistic spectrum disorder) and the self-reports of comprehensive school students. *Educational Psychology in Practice*, *16*(3), 285–296.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale, second edition (SRS-2).* Torrance, CA: Western Psychological Services.

Corbett, B. A., & Constantine, L. J. (2006). Autism and attention deficit hyperactivity disorder: assessing attention and response control with the integrated visual and auditory continuous performance test. *Child Neuropsychology*, *12*(4–5), 335–348.

Corbett, B. A., Constantine, L. J., Hendren, R., Rocke, D., & Ozonoff, S. J. (2009). Examining executive functioning in children with autism spectrum disorder, attention deficit hyperactivity disorder and typical development. *Psychiatry Research*, *166*(2–3), 210–222.

Courchesne, E. (2002). Abnormal early brain development in autism. *Molecular Psychiatry*, *7*(S2), 21–23.

Courchesne, E., Karns, C. M., Davis, H. R., Ziccardi, R., Carper, R. a, Tigue, Z. D., … Courchesne, R. Y. (2011). Unusual brain growth patterns in early life in patients with autistic disorder: an mri study. *Neurology*, *57*(24), 2111–2111.

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., de Giambattista, C., Margari, L., … de Giambattista, C. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment*, *12*, 1191–1202.

Crane, L., Adams, F., Harper, G., Welch, J., & Pellicano, E. (2018). 'Something needs to change': mental health experiences of young autistic adults in england. *Autism*, 55–59.

Crane, L., Goddard, L., & Pring, L. (2013). Autobiographical memory in adults with autism spectrum disorder: the role of depressed mood, rumination, working memory and theory of mind. *Autism*, *17*(2), 205–219.

Cribb, S., Kenny, L., & Pellicano, E. (2018). "I definitely feel more in control of my life": the perspectives of young autistic people and their parents on emerging adulthood.

Under Submission.

Croen, L. A., Zerbo, O., Qian, Y., Massolo, M. L., Rich, S., Sidney, S., & Kripke, C. C. (2015). The health status of adults on the autism spectrum. *Autism*, *19*(7), 814–823.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Cui, J., Gao, D., Chen, Y., Zou, X., & Wang, Y. (2010). Working memory in early-school-age children with asperger's syndrome. *Journal of Autism and Developmental Disorders*, *40*(8), 958–967.

Culbertson, W. C., & Zillmer, E. A. (1998). The tower of london(dx): a standardized approach to assessing executive functioning in children. *Archives of Clinical Neuropsychology*, *13*(3), 285–301.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of rdoc. *BMC Medicine*, *11*(1), 126.

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review*, *31*(6), 965–982.

D'Cruz, A.-M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, *27*(2), 152–160.

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571–578.

Damasio, A. R. (1995). Toward a neurobiology of emotion and feeling: operational concepts and hypotheses. *The Neuroscientist*, *1*(1), 19–25.

Damasio, A. R., & Maurer, R. G. (1978). A neurological model for childhood autism.

Archives of Neurology, *35*(12), 777–786.

Damiano, C. R., Mazefsky, C. A., White, S. W., & Dichter, G. S. (2014). Future directions for research in autism spectrum disorders. *Journal of Clinical Child and Adolescent Psychology*, *43*(5), 828–843.

Davids, R. C. D., Groen, Y., Berg, I. J., Tucha, O. M., & van Balkom, I. D. (2016). Executive functions in older adults with autism spectrum disorder: objective performance and subjective complaints. *Journal of Autism and Developmental Disorders*, *46*(9), 2859–2873.

Dawson, G., Meltzoff, A. N., Osterling, J., & Rinaldi, J. (1998). Neuropsychological correlates of early symptoms of autism. *Child Development*, *69*(5), 1276–1285.

Dawson, G., Munson, J. A., Estes, A. M., Osterling, J., Dawson, G., Munson, J. A., … Abbott, R. (2016). Neurocognitive function and joint attention ability in young children with autism spectrum disorder versus developmental delay. *Child Development*, *73*(2), 345–358.

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Ercument Cicek, A., … Buxbaum, J. D. (2014). Supple: synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209–215.

de Schipper, E., Lundequist, A., Coghill, D., de Vries, P. J., Granlund, M., Holtmann, M., … Bölte, S. (2015). Ability and disability in autism spectrum disorder: a systematic literature review employing the international classification of functioning, disability and health-children and youth version. *Autism Research*, *8*(6), 782–794.

de Schipper, E., Mahdi, S., de Vries, P., Granlund, M., Holtmann, M., Karande, S., … Bölte, S. (2016). Functioning and disability in autism spectrum disorder: a worldwide survey of experts. *Autism Research*, *9*(9), 959–969.

de Vries, M., & Geurts, H. M. (2012). Cognitive flexibility in asd; task switching with emotional faces. *Journal of Autism and Developmental Disorders*, *42*(12), 2558–2568.

de Vries, M., & Geurts, H. M. (2015). Influence of autism traits and executive functioning on quality of life in children with an autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(9), 2734–2743.

De Vries, M., & Geurts, H. M. (2014). Beyond individual differences: are working memory and inhibition informative specifiers within asd? *Journal of Neural Transmission*, *121*(9), 1183–1198.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001a). *Delis-Kaplan Executive Function System: Examiners manual*. Psychological Corporation.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001b). *Delis-Kaplan Executive Function System®(D-KEFS®): Examiner's Manual: flexibility of thinking, concept formation, problem solving, planning, creativity, impluse control, inhibition*. Pearson.

Delobel-Ayoub, M., Ehlinger, V., Klapouszczak, D., Maffre, T., Raynaud, J. P., Delpierre, C., & Arnaud, C. (2015). Socioeconomic disparities and prevalence of autism spectrum disorders and intellectual disability. *PLoS ONE*, *10*(11), 1–13.

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., … Guastella, A. J. (2018). Autism spectrum disorders: a meta-analysis of executive function. *Molecular Psychiatry*, *23*(5), 1198–1204.

DePape, A.-M., & Lindsay, S. (2016). Lived experiences from the perspective of individuals with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, *31*(1), 60–71.

DeSouza, J. F. X. (2002). Preparatory set associated with pro-saccades and anti-saccades in humans investigated with event-related fmri. *Journal of Neurophysiology*, *89*(2), 1016–1023.

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: a meta-analysis. *Child Development*, *85*(5), 1777–1794.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Diamond, A. (2014). Executive functions: insights into ways to help more children thrive. *Zero to Three*, *52*(2), 9–18.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855), 1387–1388.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959–964.

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, *18*, 34–48.

Dichter, G. S., & Belger, A. (2007). Social stimuli interfere with cognitive control in autism. *NeuroImage*, *35*(3), 1219–1230.

Dichter, G. S., & Belger, A. (2008). Atypical modulation of cognitive control by arousal in autism. *Psychiatry Research - Neuroimaging*, *164*(3), 185–197.

Dichter, G. S., Radonovich, K. J., Turner-Brown, L. M., Lam, K. S. L., Holtzclaw, T. N., & Bodfish, J. W. (2010). Performance of children with autism spectrum disorders on the dimension-change card sort task. *Journal of Autism and Developmental Disorders*, *40*(4), 448–456.

Dickie, V. A., Baranek, G. T., Schultz, B., Watson, L. R., & McComish, C. S. (2009). Parent reports of sensory experiences of preschool children with and without autism: a qualitative study. *American Journal of Occupational Therapy*, *63*(2), 172–181.

Dijkhuis, R. R., Ziermans, T. B., Van Rijn, S., Staal, W. G., & Swaab, H. (2017). Self-regulation and quality of life in high-functioning young adults with autism. *Autism*, *21*(7), 896–906.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259–268.

Duncan, J. (1986). Disorganisation of behaviour after frontal lobe damage. *Cognitive Neuropsychology*, *3*(3), 271–290.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: unity and diversity of function. *Cognitive Psychology*, *14*(5), 713–741.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483.

Durrleman, S., Franck, J., Stephanie, D., & Julie, F. (2015). Exploring links between language and cognition in autism spectrum disorders: complement sentences, false belief, and executive functioning. *Journal of Communication Disorders*, *54*, 15–31.

Ecker, C. (2017). The neuroanatomy of autism spectrum disorder: an overview of structural neuroimaging findings and their translatability to the clinical setting. *Autism*, *21*(1), 18–28.

Edgin, J. O., & Pennington, B. F. (2005). Spatial cognition in autism spectrum disorders: superior, impaired, or just intact? *Journal of Autism and Developmental Disorders*, *35*(6), 729–745.

Eisenberg, I. W., Bissett, P. G., Canning, J. R., Dallery, J., Enkavi, A. Z., Whitfield-Gabrieli, S., … Poldrack, R. A. (2017). Applying novel technologies and methods to inform the ontology of self-regulation. *Behaviour Research and Therapy*, *101*(April 2017), 46–57.

Elliott, R., Fischer, C., & Rennie, D. (1999). Evolving guidelines for publication of pualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, *38*, 215–229.

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., … Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160–179.

Englund, J. A., Decker, S. L., Allen, R. A., & Roberts, A. M. (2014). Common cognitive deficits in children with attention-deficit/hyperactivity disorder and autism: working memory and visual-motor integration. *Journal of Psychoeducational Assessment*, *32*(2), 95–106.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception and Psychophysics*, *16*(1), 143–149.

Eskes, G. A., Bryson, S. E., & McCormick, T. A. (1990). Comprehension of concrete and abstract words in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 61–73.

Evans, J. J., Chua, S. E., Mckenna, P. J., & Wilson, B. A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine*, *27*(3), 635–646.

Faja, S., & Dawson, G. (2013). Reduced delay of gratification and effortful control among young children with autism spectrum disorders. *Autism*, *19*(1), 91–101.

Faja, S., & Dawson, G. (2014). Performance on the dimensional change card sort and backward digit span by young children with autism without intellectual disability. *Child Neuropsychology*, *20*(6), 692–699.

Faja, S., Dawson, G., Sullivan, K., Meltzoff, A. N., Estes, A., & Bernier, R. (2016). Executive function predicts the development of play skills for verbal preschoolers with autism spectrum disorders. *Autism Research*, *9*(12), 1274–1284.

Fein, E. (2016). Our circuits, ourselves: what the autism spectrum can tell us about the research domain criteria project (rdoc) and the neurogenetic transformation of diagnosis. *BioSocieties*, *11*(2), 175–198.

Fine, C. (2014). His brain, her brain? *Science*, *346*(6212), 915–916.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327–358.

Fletcher-Watson, S., Adams, J., Brook, K., Charman, T., Crane, L., Cusack, J., … Pellicano, E. (2018). Making the future together: shaping autism research through meaningful participation. *Autism*, 136236131878672.

Fletcher, A. J. (2017). Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology*, *20*(2), 181–194.

Folstein, S., & Rutter, M. (1977). Infantile autism: a genetic study of 21 twin pairs. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *18*(4), 297–321.

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, *65*(6), 591–598.

Fornasari, L., Chittaro, L., Ieronutti, L., Cottini, L., Dassi, S., Cremaschi, S., … Brambilla, P. (2013). Navigation and exploration of an urban virtual environment by children with autism spectrum disorder compared to children with typical development. *Research in Autism Spectrum Disorders*, *7*(8), 956–965.

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *Journal of Autism and Developmental Disorders*, *40*(10), 1227–1240.

Franzen, M. D., & Arnett, P. A. (1997). The validity of neuropsychological assessment procedures. In H. Reese & M. Franzen (Eds.), *Biological and neuropsychological mechanisms: Life-span developmental psychology* (pp. 51–69). Mahwah, NJ: Lawrence Erlbaum

Associates, Inc.

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 91–112). Delray Beach, FL: St Lucie Press, Inc.

Frith, C. (2004). Is autism a disconnection disorder? *Lancet Neurology*, *3*(10), 577.

Frith, U., & Happé, F. G. E. (1999). Theory of mind and self consciousness: what is it like to be autistic? *Mind & Language*, *14*(1), 1–22.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: autism. *Trends in Neurosciences*, *14*(10), 433–438.

Gabig, C. S. (2008). Verbal working memory and story retelling in school-age children with autism. *Language Speech and Hearing Services in Schools*, *39*(4), 498.

Galpin, J., Barratt, P., Ashcroft, E., Greathead, S., Kenny, L., & Pellicano, E. (2017). 'The dots just don't join up': understanding the support needs of families of children on the autism spectrum. *Autism*, 136236131668798.

Gan, S.-M., Tung, L.-C., Yeh, C.-H., Chang, H.-Y., & Wang, C.-H. (2014). The icf-cy-based structural equation model of factors associated with participation in children with autism. *Developmental Neurorehabilitation*, *17*(1), 24–33.

García-Villamisar, D., & Della Sala, S. (2002). Dual-task performance in adults with autism. *Cognitive Neuropsychiatry*, *7*(1), 63–74.

Garin, O., Ayuso-Mateos, J. L., Almansa, J., Nieto, M., Chatterji, S., Vilagut, G., … MHADIE consortium. (2010). Validation of the "world health organization disability assessment schedule, whodas-2" in patients with chronic diseases. *Health and Quality of Life Outcomes*, *8*(1), 51.

Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., … Buxbaum,

J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–5.

Gelman, A., & Loken, E. (2014). The garden of forking paths: why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Psychological Bulletin*, *140*(5), 1272–1280.

Gernsbacher, M. A. (2017). Editorial perspective: the use of person-first language in scholarly writing may accentuate stigma. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(7), 859–861.

Gernsbacher, M. A., Dawson, M., & Goldsmith, H. H. (2005). Three reasons not to believe in an autism epidemic. *Current Directions in Psychological Science*, *14*(2), 55–58.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 3.5-7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129–153.

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. *Current Opinion in Neurobiology*.

Geurts, H. M., Begeer, S., & Stockmann, L. (2009). Brief report: inhibitory control of socially relevant stimuli in children with high functioning autism. *Journal of Autism and Developmental Disorders*, *39*(11), 1603–1607.

Geurts, H. M., Corbett, B. A., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in Cognitive Sciences*, *13*(2), 74–82.

Geurts, H. M., Luman, M., & Van Meel, C. S. (2008). What's in a game: the effect of social motivation on interference control in boys with adhd and autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *49*(8), 848–857.

Geurts, H. M., van den Bergh, S. F. W. M., & Ruzzano, L. (2014). Prepotent response inhibition and interference control in autism spectrum disorders: two meta-analyses. *Autism Research*, *7*(4), 407–420.

Geurts, H. M., Verté, S., Oosterlaan, J., Roeyers, H., & Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(4), 836–854.

Geurts, H. M., & Vissers, M. E. (2012). Elderly with autism: executive functions and memory. *Journal of Autism and Developmental Disorders*, *42*(5), 665–675.

Gillberg, C. I., & Steffenburg, S. (1987). Outcome and prognostic factors in infantile autism and similar conditions: a population-based study of 46 cases followed through puberty. *Journal of Autism and Developmental Disorders*, *17*(2), 273–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3610999

Gillespie-Lynch, K., Kapp, S. K., Brooks, P. J., Pickens, J., & Schwartzman, B. (2017). Whose expertise is it? evidence for autistic adults as critical autism experts. *Frontiers in Psychology*, *8*(MAR), 1–14.

Gilotty, L., Kenworthy, L. E., Wagner, A. E., Sirian, L., Black, D. O., & Wagner, A. E. (2002). Adaptive skills and executive function in autism spectrum disorders. *Child Neuropsychology (Neuropsychology, Development and Cognition: Section C)*, *8*(4), 241–248.

Gioia, G. A., Guy, S. C., Isquith, P. K., & Kenworthy, L. E. (1996). *Behavior rating inventory of executive function*. Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. E. (2015). *Behavior Rating Inventory of Executive Function; Professional Manual, Second edition* (2nd ed.). Florida: PAR.

Glass, G. V, Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237–288.

Goldberg, M. C., Lasker, A. G., Zee, D. S., Garth, E., Tien, A. Y., & Landa, R. J. (2002). Deficits in the initiation of eye movements in the absence of a visual target in adolescents with high functioning autism. *Neuropsychologia*, *40*(12), 2039–2049.

Goldberg, M. C., Mostofsky, S. H., Cutting, L. E., Mahone, E. M., Astor, B. C., Denckla, M. B., & Landa, R. J. (2005). Subtle executive impairment in children with autism and children with adhd. *Journal of Autism and Developmental Disorders*, *35*(3), 279–293.

Goldstein, G., Johnson, C. R., & Minshew, N. J. (2001). Attentional processes in autism. *Journal of Autism and Developmental Disorders*, *31*(4), 433–440.

Goldstein, G., Minshew, N. J., Allen, D. N., & Seaton, B. E. (2002). High-functioning autism and schizophrenia: a comparison of an early and late onset neurodevelopmental disorder. *Archives of Clinical Neuropsychology*, *17*(5), 461–475.

Gonzalez-Gadea, M. L. M. L., Baez, S., Torralva, T., Castellanos, F. X., Rattazzi, A., Bein, V., … Ibanez, A. A. (2013). Cognitive variability in adults with adhd and as: disentangling the roles of executive functions and social cognition. *Research in Developmental Disabilities*, *34*(2), 817–830.

Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ados scores. *Pediatrics*, *130*(5), e1278–e1284.

Grace, J., & Malloy, P. F. (2001). *Frontal systems behavior scale: FrSBe*. Lutz, FL: Psychological Assessment Resources.

Grant, D., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411.

Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science*, *7*(2), 99–108.

Grewe, P., Lahr, D., Kohsik, A., Dyck, E., Markowitsch, H. J., Bien, C. G., … Piefke, M. (2014). Real-life memory and spatial navigation in patients with focal epilepsy: ecological validity of a virtual reality supermarket task. *Epilepsy and Behavior*, *31*, 57–66.

Griebling, J., Minshew, N. J., Bodner, K. E., Libove, R., Bansal, R., Konasale, P., … Hardan, A. Y. (2010). Dorsolateral prefrontal cortex magnetic resonance imaging measurements and cognitive performance in autism. *Journal of Child Neurology*, *25*(7), 856–63.

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development*, *70*(4), 817–832.

Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (asds): an opportunity for identifying asd subtypes. *Molecular Autism*, *4*(1), 12.

Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Haigh, S. M., Gupta, A., Barb, S. M., Glass, S. A. F., Minshew, N. J., Dinstein, I., … Behrmann, M. (2016). Differential sensory fmri signatures in autism and schizophrenia: analysis of amplitude and trial-to-trial variability. *Schizophrenia Research*, *175*(1–3), 12–19.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296.

Ham, H. S., Bartolo, A., Corley, M., Rajendran, G., Szabo, A., & Swanson, S. (2011). Exploring the relationship between gestural recognition and imitation: evidence of dyspraxia in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *41*(1), 1–12.

Hammond, K. R. (1998). Ecological validity: then and now. Retrieved February 1, 2018, from https://www.albany.edu/cpr/brunswik/notes/essay2.html

Hansen, S. N., Schendel, D. E., & Parner, E. T. (2015). Explaining the increase in the prevalence of autism spectrum disorders. *JAMA Pediatrics*, *169*(1), 56.

Hanson, L. K., & Atance, C. M. (2014). Brief report: episodic foresight in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(3), 674–684.

Happé, F. G. E. (1994). *Autism: An introduction to psychological theory*. London: UCL Press.

Happé, F. G. E. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. a research note. *Journal of Child Psychology and Psychiatry*, *37*(7), 873–877.

Happé, F. G. E., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. *Brain and Cognition*, *61*(1), 25–39.

Happé, F. G. E., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25.

Happé, F. G. E., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*(10), 1218–1220.

Harris, J. M., Best, C. S., Moffat, V. J., Spencer, M. D., Philip, R. C. M., Power, M. J., & Johnstone, E. C. (2008). Autistic traits and cognitive performance in young people with mild intellectual impairment. *Journal of Autism and Developmental Disorders*, *38*(7), 1241–1249.

Hay, I., & Winn, S. (2005). Students with asperger's syndrome in an inclusive secondary school environment: teachers', parents', and students' perspectives. *Australasian Journal of Special Education*, *29*(2), 140–154.

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in*

Cognitive Sciences, *9*(4), 188–194.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. In *Journal of Vision* (Vol. 3, p. 6).

Hebron, J., & Humphrey, N. (2014). Mental health difficulties among young people on the autistic spectrum in mainstream secondary schools: a comparative study. *Journal of Research in Special Educational Needs*, *14*(1), 22–32.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1–21.

Heinrichs, R. W. (1990). Current and emergent applications of neuropsychological assessment: problems of validity and utility. *Professional Psychology: Research and Practice*, *21*(3), 171–176.

Henderson, H., Schwartz, C., Mundy, P., Burnette, C., Sutton, S., Zahka, N., & Pradella, A. (2006). Response monitoring, the error-related negativity, and differences in social behavior in autism. *Brain and Cognition*, *61*(1), 96–109.

Hill, E. L. (2004a). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, *24*(2), 189–233.

Hill, E. L. (2004b). Executive dysfunction in autism. *Trends in Cognitive Sciences*, *8*(1), 26–32.

Hill, E. L., & Bird, C. M. (2006). Executive processes in asperger syndrome: patterns of performance in a multiple case series. *Neuropsychologia*, *44*(14), 2822–2835.

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.

Hollocks, M. J., Jones, C. R. G., Pickles, A., Baird, G., Happé, F. G. E., Charman, T., & Simonoff, E. (2014). The association between social cognition and executive functioning and symptoms of anxiety and depression in adolescents with autism spectrum disorders. *Autism Research*, *7*(2), 216–228.

Horwitz, J. E., & McCaffrey, R. J. (2008). Effects of a third party observer and anxiety on tests of executive function. *Archives of Clinical Neuropsychology*, *23*(4), 409–417.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry*, *45*(2), 212–229.

Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal iq as children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*(6), 572–581.

Howlin, P., Savage, S., Moss, P., Tempier, A., & Rutter, M. (2014). Cognitive and language skills in adults with autism: a 40-year follow-up. *Journal of Child Psychology and Psychiatry*, *55*(1), 49–58.

Hu, L., Zang, Y. L., & Li, N. (2012). The applicability of whodas 2.0 in adolescents in china. *Journal of Clinical Nursing*, *21*(17–18), 2438–2451.

Hughes, C. (1998). Executive function in preschoolers: links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*(2), 233–253.

Hughes, C., Leboyer, M., Bouvard, M., C., H., M., L., & Bouvard, M. (1997). Executive function in parents of children with autism. *Psychological Medicine*, *27*(1), 209–220.

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with mental disengagement from an object: its implications for theories of autism. *Developmental Psychology*, *29*(3), 498–510.

Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia*, *32*(4), 477–492.

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*, *47*(8), 1–16.

Humphrey, N., & Lewis, S. (2008). "Make me normal": the views and experiences of pupils on the autistic spectrum in mainstream secondary schools. *Autism*, *12*(1), 23–46.

Hutton, S. B., Puri, B. K., Duncan, L. ., Robbins, T. W., Barnes, T. R. E., & Joyce, E. M. (2017). Executive function in first-episode schizophrenia. *Psychological Medicine*, *28*(2), 463–473.

Hyde, M., Scott-Slade, M., Scott-Slade, H., Hornberger, M., Spiers, H. J., Dalton, R., … Bohbot, V. (2016). Sea hero quest: the world's first mobile game where anyone can help scientists fight dementia.

Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology*, *6*(1), 155–179.

Iarocci, G., & Burack, J. A. (2004). Intact covert orienting to peripheral cues among children with autism. *Journal of Autism and Developmental Disorders*, *34*(3), 257–264.

Idring, S., Rai, D., Dal, H., Dalman, C., Sturm, H., Zander, E., … Magnusson, C. (2012). Autism spectrum disorders in the stockholm youth cohort: design, prevalence and validity. *PLoS ONE*, *7*(7), e41280.

Insel, T. R., Cuthbert, B. N., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research domain criteria (rdoc): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751.

Jack, A. I., & Roepstorff, A. (2002). Introspection and cognitive brain mapping: from stimulus-response to script-report. *Trends in Cognitive Sciences*, *6*(8), 333–339.

Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., … Eichler, E. E. (2014). A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *American Journal of Human Genetics*, *94*(3), 415–425.

Jahromi, L. B., Bryce, C. I., & Swanson, J. (2013). The importance of self-regulation for the school and peer engagement of children with high-functioning autism. *Research in Autism Spectrum Disorders*, *7*(2), 235–246.

Jansari, A. S., Devlin, A., Agnew, R., Akesson, K., Murphy, L., & Leadbetter, T. (2014). Ecological assessment of executive functions: a new virtual reality paradigm. *Brain Impairment*, *15*(2), 71–87.

Jarrold, C., & Brock, J. (2004). To match or not to match? methodological issues in autism related research. *Journal of Autism and Developmental Disorders*, *34*(1), 81–86.

Jenkins, C. S. (2010). What is ontological realism? *Philosophy Compass*, *5*(10), 880–890.

Johnston, K., Madden, A. K., Bramham, J., & Russell, A. J. (2011). Response inhibition in adults with autism spectrum disorder compared to attention deficit/hyperactivity disorder. *Journal of Autism and Developmental Disorders*, *41*(7), 903–912.

Jones, C. R. G., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J. S., Tregay, J., … Charman, T. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, *11*(1), 95–109.

Jones, R. S. P., Quigney, C., & Huws, J. C. (2003). First-hand accounts of sensory perceptual experiences in autism: a qualitative analysis. *Journal of Intellectual and Developmental Disability*, *28*(2), 112–121.

Joseph, R. M., Mcgrath, L. M., & Tager-Flusberg, H. (2013). Executive dysfunction and its relation to language ability in verbal school-age children with autism. *Developmental Neuropsychology*, *27*(3), 361–378.

Joseph, R. M., Steele, S. D., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: failure to use verbal mediation in the service of working memory? *Neuropsychologia*, *43*(10), 1400–1411.

Joseph, R. M., Tager-Flusberg, H., Lord, C., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *43*(6), 807–821.

Joshi, G., Petty, C., Wozniak, J., Henin, A., Fried, R., Galdo, M., … Biederman, J. (2010). The heavy burden of psychiatric comorbidity in youth with autism spectrum disorders: a large comparative study of a psychiatrically referred population. *Journal of Autism and Developmental Disorders*, *40*(11), 1361–1370.

Joshi, G., Wozniak, J., Petty, C., Martelon, M. K., Fried, R., Bolfek, A., … Biederman, J. (2013). Psychiatric comorbidity and functioning in a clinically referred population of adults with autism spectrum disorders: a comparative study. *Journal of Autism and Developmental Disorders*, *43*(6), 1314–1325.

Kable, J. W., Caulfield, M. K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., … Lerman, C. (2017). No effect of commercial cognitive training on brain activity, choice behavior, or cognitive performance. *The Journal of Neuroscience*, *37*(31), 7390–7402.

Kalbfleisch, M. L., & Loughan, A. R. (2012). Impact of iq discrepancy on executive function in high-functioning autism: insight into twice exceptionality. *Journal of Autism and Developmental Disorders*, *42*(3), 390–400.

Kamp-Becker, I., Schröder, J., Remschmidt, H., Bachmann, C. J., & Schroder, J. (2010). Health-related quality of life in adolescents and young adults with high functioning autism-spectrum disorder. *GMS PsychoSocialMedicine*, *7*, 1–10.

Kana, R. K., Keller, T. A., Minshew, N. J., & Just, M. A. (2007). Inhibitory control in high-functioning autism: decreased activation and underconnectivity in inhibition networks. *Biological Psychiatry*, *62*(3), 198–206.

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when. *Psychological Science*, *18*(7), 614–621.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615–622.

Kanner, L. (1943). Autistic disturbance of affective contact. *Nervous Child*, *2*, 217–250.

Kanner, L. (1949). Problems of nosology and psychodynamics of early infantile autism. *American Journal of Orthopsychiatry*, *19*(3), 416.

Kapp, S. K., Gillespie-Lynch, K., Sherman, L. E., & Hutman, T. (2013). Deficit, difference, or both? autism and neurodiversity. *Developmental Psychology*, *49*(1), 59–71.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2*(10), 389–398.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Developmental Psychology*, *45*(1), 56–63.

Kaufmann, L., Zotter, S., Pixner, S., Starke, M., Haberlandt, E., Steinmayr-Gensluckner, M., … Marksteiner, J. (2013). Brief report: cantab performance and brain structure in pediatric patients with asperger syndrome. *Journal of Autism and Developmental Disorders*, *43*(6), 1483–1490.

Keary, C. J., Minshew, N. J., Bansal, R., Goradia, D., Fedorov, S., Keshavan, M. S., & Hardan, A. Y. (2009). Corpus callosum volume and neurocognition in autism. *Journal of Autism and Developmental Disorders*, *39*(6), 834–841.

Kenny, L., Cribb, S., & Pellicano, E. (2018). Childhood theory of mind, planning and cognitive flexibility predict later behavioural outcomes in autistic adolescents. *Manuscript in Submission*.

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.

Kenworthy, L. E., Anthony, L. G., Naiman, D. Q., Cannon, L., Wills, M. C., Luong-Tran, C., … Wallace, G. L. (2014). Randomized controlled effectiveness trial of executive function intervention for children on the autism spectrum. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(4), 374–383.

Kenworthy, L. E., Black, D. O., & Wallace, G. L. (2005). Disorganization: the forgotten executive dysfunction in high- functioning autism (hfa) spectrum disorders. *Developmental Neuropsychology*, *28*(3), 809–827.

Kenworthy, L. E., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review*, *18*(4), 320–338.

Khundrakpam, B. S., Lewis, J. D., Kostopoulos, P., Carbonell, F., & Evans, A. C. (2017). Cortical thickness abnormalities in autism spectrum disorders through late childhood, adolescence , and adulthood: a large-scale mri study. *Cerebral Cortex*, *27*(3), 1721–1731.

Kibby, M. Y., Schmitter-Edgecombe, M., & Long, C. J. (1998). Ecological validity of neuropsychological tests: focus on the california verbal learning test and the wisconsin card sorting test. *Archives of Clinical Neuropsychology*, *13*(6), 523–34.

Kiep, M., & Spek, A. A. (2017). Executive functioning in men and women with an autism spectrum disorder. *Autism Research*, *10*(5), 940–948.

Kilinçaslan, A., Mukaddes, N. M., Küçükyazici, G. S., & Gürvit, H. (2010). Assessment of executive/attentional performance in asperger's disorder. *Turkish Journal of Psychiatry*, *21*(4), 289–299.

Kim, Y. S., Leventhal, B. L., Koh, Y.-J., Fombonne, E., Laska, E., Lim, E.-C., … Grinker, R. R. (2011). Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry*, *168*(9), 904–912.

Kimhi, Y., Shoam-Kugelmas, D., Agam Ben-Artzi, G., Ben-Moshe, I., & Bauminger-Zviely, N. (2014). Theory of mind and executive function in preschoolers with typical development versus intellectually able preschoolers with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(9), 2341–2354.

King, K. M., Patock-Peckham, J. A., Dager, A. D., Thimm, K., & Gates, J. R. (2014). On the mismeasurement of impulsivity: trait, behavioral, and neural models in alcohol research among adolescents and young adults. *Current Addiction Reports*, *1*(1), 19–32.

Kipp, M. (2012). Multimedia annotation, querying, and analysis in anvil. *Multimedia Information Extraction: Advances in Video, Audio, and Imagery Analysis for Search, Data Mining, Surveillance, and Authoring*, 351–367.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358.

Klin, A., Saulnier, C. A., Sparrow, S. S., Cicchetti, D. V., Volkmar, F. R., & Lord, C. (2007). Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: the vineland and the ados. *Journal of Autism and Developmental Disorders*, *37*(4), 748–759.

Klinger, E., Ph, D., Chemin, I., Lebreton, S., Psy, D., & Marié, R. (2006). Virtual action planning in parkinson’s disease: a control study. *Cyberpsychology and Behavior*, *9*(3), 342–347.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY–Second Edition (NEPSY-II)*. San Antonio: The Psychological Corporation.

Koshino, H., Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2008). FMRI investigation of working memory for faces in autism: visual coding and underconnectivity with frontal areas. *Cerebral Cortex*, *18*(2), 289–300.

Kouklari, E.-C., Tsermentseli, S., & Monks, C. P. (2018). Everyday executive function and adaptive skills in children and adolescents with autism spectrum disorder: cross-sectional developmental trajectories. *Autism & Developmental Language Impairments*, *3*, 239694151880077.

Kramer, J. H., Mungas, D. M., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., … Widmeyer, M. (2013). *Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research (EXAMINER), User Manual, version 3.6*. NIH-EXAMINER team.

Kramer, J. H., Mungas, D. M., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., … Widmeyer, M. (2014). NIH examiner: conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society*, *20*(1), 11–9.

Kruglanski, A. W. (1975). The human subject in the psychology experiment: fact and artifact. In *Advances in experimental social psychology* (Vol. 8, pp. 101–147). Elsevier.

Kvavilashvili, L., & Ellis, J. (2004). Ecological validity and the real-life/laboratory controversy in memory research: a critical (and historical) review. *History & Philosophy of Psychology*, *6*(1), 59–80.

Lai, E. C. L., Lau, Z., Lui, S. S. Y., Lok, E., Tam, V., Chan, Q., … Cheung, E. F. C. (2017). Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. *Autism Research*, *10*(5), 911–939.

Lai, M.-C., Lombardo, M. V., Pasco, G., Ruigrok, A. N., Wheelwright, S. J., Sadek, S. A., …

MRC AIMS Consortium. (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PloS One*, *6*(6), e20835.

Lai, M.-C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., … MRC AIMS Consortium. (2016). Quantifying and exploring camouflaging in men and women with autism. *Autism*, 1–13.

Landa, R. J., & Goldberg, M. C. (2005). Language, social, and executive functions in high functioning autism: a continuum of performance. *Journal of Autism and Developmental Disorders*, *35*(5), 557–573.

Landry, O., & Al-Taie, S. (2016). A meta-analysis of the wisconsin card sort task in autism. *Journal of Autism and Developmental Disorders*, *46*(4), 1220–1235.

Lange, N., Travers, B. G., Bigler, E. D., Prigge, M. B. D., Froehlich, A. L., Nielsen, J. A., … Lainhart, J. E. (2015). Longitudinal volumetric brain changes in autism spectrum disorder ages 6-35 years. *Autism Research*, *8*(1), 82–93.

Langen, M., Leemans, A., Johnston, P., Ecker, C., Daly, E., Murphy, C. M., … Murphy, D. G. M. (2012). Fronto-striatal circuitry and inhibitory control in autism: findings from diffusion tensor imaging tractography. *Cortex*, *48*(2), 183–193.

Lawson, R. A., Papadakis, A. A., Higginson, C. I., Barnett, J. E., Wills, M. C., Strang, J. F., … Kenworthy, L. E. (2015). Everyday executive function impairments predict comorbid psychopathology in autism spectrum and attention deficit hyperactivity disorders. *Neuropsychology*, *29*(3), 445–453.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*(May), 1–10.

Lee, P. S., Yerys, B. E., Della Rosa, A., Foss-Feig, J., Barnes, K. A., James, J. D., … Kenworthy, L. E. (2009). Functional connectivity of the inferior frontal cortex changes with age in children with autism spectrum disorders: a fcmri study of response inhibition. *Cerebral Cortex*, *19*(8), 1787–1794.

Lemon, J. M., Gargaro, B., Enticott, P. G., & Rinehart, N. J. (2011). Brief report: executive functioning in autism spectrum disorders: a gender comparison of response inhibition. *Journal of Autism and Developmental Disorders*, *41*(3), 352–356.

Leung, R. C., & Zakzanis, K. K. (2014). Brief report: cognitive flexibility in autism spectrum disorders: a quantitative review. *Journal of Autism and Developmental Disorders*, *44*(10), 2628–2645.

Lever, A. G., & Geurts, H. M. (2016). Psychiatric co-occurring symptoms and disorders in young, middle-aged, and older adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(6), 1916–1930.

Lever, A. G., Werkle-Bergner, M., Brandmaier, A. M., Ridderinkhof, K. R., & Geurts, H. M. (2015). Atypical working memory decline across the adult lifespan in autism spectrum disorder? *Journal of Abnormal Psychology*, *124*(4), 1014–1026.

Leyfer, O. T., Folstein, S. E., Bacalman, S., Davis, N. O., Dinh, E., Morgan, J., … Lainhart, J. E. (2006). Comorbid psychiatric disorders in children with autism: interview development and rates of disorders. *Journal of Autism and Developmental Disorders*, *36*(7), 849–861.

Lezak, M. D. (1983). Verbal functions. In *Neuropsychological Assessment* (pp. 312–341). New York, NY: Oxford University Press.

Liberatos, P., Link, B. G., & Kelsey, J. L. (1987). The measurement of social class in epidemiology. *Epidemiologic Reviews*, *10*(1), 87–121.

Limoges, É., Bolduc, C., Berthiaume, C., Mottron, L., & Godbout, R. (2013). Relationship between poor sleep and daytime cognitive performance in young adults with autism. *Research in Developmental Disabilities*, *34*(4), 1322–1335.

Lisiecka, D. M., Holt, R. J., Tait, R., Ford, M., Lai, M.-C., Chura, L. R., … Suckling, J. (2015). Developmental white matter microstructure in autism phenotype and corresponding endophenotype during adolescence. *Translational Psychiatry*, *5*(3), e529–e529.

Liss, M., Fein, D. A., Allen, D., Dunn, M., Feinstein, C., Morris, R., … Rapin, I. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(2), 261–270.

Livingston, L. A., & Happé, F. G. E. (2017). Conceptualising compensation in neurodevelopmental disorders: reflections from autism spectrum disorder. *Neuroscience and Biobehavioral Reviews Biobehavioral Reviews*.

Llorente, A. M., Williams, J., Satz, P., & D'Elia, L. F. (2003). *Children's color trails test (CCTT)*. Odessa, Florida: Psychological Assessment Resources.

Logan, G. D. (1994). On the ability to inhibit thought and action: a users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego, CA, US: Academic Press.

Logie, R. H. (2016). Retiring the central executive. *Quarterly Journal of Experimental Psychology*, *69*(10), 2093–2109.

London, E. B. (2014). Categorical diagnosis: a fatal flaw for autism research? *Trends in Neurosciences*, *37*(12), 683–686.

Long, C. J., & Kibby, M. Y. (1995). Ecological validity of neuropsychological tests: a look at neuropsychology's past and the impact that ecological issues may have on its future. *Advances in Medical Psychotherapy*, *8*(5).

Loomes, R., Hull, L., & Mandy, W. (2017). What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(6), 466–474.

Lopez, B. R., Lincoln, A. J., Ozonoff, S. J., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of autistic disorder. *Journal of Autism and Developmental Disorders*, *35*(4), 445–460.

Lord, C., Petkova, E., Hus, V., Gan, W., & Lu, F. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatryeneral*, *69*(3), 306–313.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule – 2nd edition (ADOS-2)*. Los Angeles, CA: Western Psychological Corporation.

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*(5), 659–685.

Losh, M., Adolphs, R., Poe, M. D., Couture, S., Penn, D., Baranek, G. T., & Piven, J. (2009). Neuropsychological profile of autism and the broad autism phenotype. *Archives of General Psychiatry*, *66*(5), 518.

Low, J., Goddard, E., & Melser, J. (2009). Generativity and imagination in autism spectrum disorder: evidence from individual differences in children's impossible entity drawings. *British Journal of Developmental Psychology*, *27*(2), 425–444.

Ludlow, A., Skelly, C., & Rohleder, P. (2012). Challenges faced by parents of children diagnosed with autism spectrum disorder. *Journal of Health Psychology*, *17*(5), 702–711.

Lugnegård, T., Hallerbäck, M. U., & Gillberg, C. (2011). Psychiatric comorbidity in young adults with a clinical diagnosis of asperger syndrome. *Research in Developmental Disabilities*, *32*(5), 1910–1917.

Lundström, S., Chang, Z., Råstam, M., Gillberg, C., Larsson, H., Anckarsäter, H., &

Lichtenstein, P. (2012). Autism spectrum disorders and autistic like traits. *Archive of General Psychiatry*, *69*(1), 46–52.

Luria, A. R. (1966). *Higher cortical functions in man*. Oxford, England: Basic Books.

Lynch, C. J., Breeden, A. L., You, X., Ludlum, R., Gaillard, W. D., Kenworthy, L. E., & Vaidya, C. J. (2017). Executive dysfunction in autism spectrum disorder is associated with a failure to modulate frontoparietal-insular hub architecture. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(6), 537–545.

Mackinlay, R., Charman, T., & Karmiloff-Smith, A. (2006). High functioning children with autism spectrum disorder: a novel test of multitasking. *Brain and Cognition*, *61*(1), 14–24.

Mahdi, S., Viljoen, M., Yee, T., Selb, M., Singhal, N., Almodayfer, O., … Bölte, S. (2018). An international qualitative study of functioning in autism spectrum disorder using the world health organization international classification of functioning, disability and health framework. *Autism Research*, *11*(3), 463–475.

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: coming out of the office. *Brain Injury*, *18*(11), 1067–1081.

Mandell, D. (2011). Editorial. *Autism*, *15*(3), 259–261.

Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. H. (2012). 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.

Mandy, W., & Lai, M.-C. (2016). Annual research review: the role of the environment in the developmental psychopathology of autism spectrum condition. *Journal of Child Psychology and Psychiatry*, *57*(3), 271–292.

Mangiafico, S. (2016). *Summary and analysis of extension program evaluation in* R (Vol. 442). New

Brunswick, NJ: Rutgers.

Mannion, A., & Leader, G. (2013). Comorbidity in autism spectrum disorder: a literature review. *Research in Autism Spectrum Disorders*, *7*(12), 1595–1616.

Marks, S. U., Schrader, C., Longaker, T., & Levine, M. (2000). Portraits of three adolescent students with asperger's syndrome: personal stories and how they can inform practice. *Research and Practice for Persons with Severe Disabilities*, *25*(1), 3–17.

Mason, J., & Scior, K. (2004). "Diagnostic overshadowing" amongst clinicians working with people with intellectual disabilities in the uk. *Journal of Applied Research in Intellectual Disabilities*, *17*(2), 85–90.

Mazefsky, C. A., Oswald, D., Day, T. N., Eack, S. M., Minshew, N. J., & Lainhart, J. E. (2012). ASD, a psychiatric disorder, or both? psychiatric diagnoses in adolescents with high-functioning asd. *Journal of Clinical Child and Adolescent Psychology*, *41*(4), 516–523.

Mazzone, L., Ruta, L., & Reale, L. (2012). Psychiatric comorbidities in asperger syndrome and high functioning autism: diagnostic challenges. *Annals of General Psychiatry*, *11*(16), 1–13.

McCrimmon, A. W., Schwean, V. L., Saklofske, D. H., Montgomery, J. M., & Brady, D. I. (2012). Executive functions in asperger's syndrome: an empirical investigation of verbal and nonverbal skills. *Research in Autism Spectrum Disorders*, *6*(1), 224–233.

McEvoy, R. E., Rogers, S. J., & Pennington, B. F. (1993). Executive function and social communication deficits in young autistic children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *34*(4), 563–578.

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Della Sala, S., Milne, A. B., … Callender, J. S. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence-*, *10*(4), 375–383.

McLaughlin, S., & Rafferty, H. (2014). Me and "it": seven young people given a diagnosis of asperger's syndrome. *Educational and Child Psychology*, *31*(1), 63–78.

McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: an experience-sampling study of mind wandering across controlled and ecological contexts. *Psychonomic Bulletin and Review*, *16*(5), 857–863.

Medeiros, K., & Winsler, A. (2014). Parent-child gesture use during problem solving in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(8), 1946–1958.

Miller, E. K., & Cohen, J. D. (2001). An integtrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked eeg. *Frontiers in Psychology*, *2*(MAR), 1–12.

Milton, D. E. (2014). Autistic expertise: a critical reflection on the production of knowledge in autism studies. *Autism*, *18*(7), 794–802.

Minshew, N. J., Meyer, J., & Goldstein, G. (2002). Abstract reasoning in autism: a disassociation between concept formation and concept identification. *Neuropsychology*, *16*(3), 327–334.

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competancies predicited by pre-school delay of gratification. *Journal of Personality and Social Psychology*, *54*(4), 687–696.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Modabbernia, A., Velthorst, E., & Reichenberg, A. (2017). Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. *Molecular Autism*, *8*(1), 13.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., … Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698.

Morady, K., & Humphreys, G. W. (2011). Multiple task demands in action disorganization syndrome. *Neurocase*, *17*(5), 461–472.

Mosconi, M. W., Kay, M., D'Cruz, A. M., Seidenfeld, A., Guter, S., Stanford, L. D., & Sweeney, J. A. (2009). Impaired inhibitory control is associated with higher-order repetitive behaviors in autism spectrum disorders. *Psychological Medicine*, *39*(9), 1559–1566.

Mostert-Kerckhoffs, M. A. L., Staal, W. G., Houben, R. H., & de Jonge, M. V. (2015). Stop and change: inhibition and flexibility skills are related to repetitive behavior in children and young adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *45*(10), 3148–3158.

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131–148). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Mottron, L., Dawson, M., Bertone, A., & Wang, L. (2007). Cognitive versatility in autism cannot be reduced to a deficit. *Cognitive Neuropsychology*, *24*(5), 578–580.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., Burack, J. J. A., Soulieres, I., … Burack, J. J. A. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43.

Murray, D., Lesser, M., & Lawson, W. (2005). Attention, monotropism and the diagnostic criteria for autism. *Autism*, *9*(2), 139–156.

Nakahachi, T., Iwase, M., Takahashi, H., Honaga, E., Sekiyama, R., Ukai, S., … Takeda, M. (2006). Discrepancy of performance among working memory-related tasks in autism spectrum disorders was caused by task characteristics, apart from working memory, which could interfere with task execution. *Psychiatry and Clinical Neurosciences*, *60*(3), 312–318.

Narzisi, A., Muratori, F., Calderoni, S., Fabbro, F., & Urgesi, C. (2013). Neuropsychological profile in high functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(8), 1895–1909.

Neil, L., Olsson, N. C., & Pellicano, E. (2016). The relationship between intolerance of uncertainty, sensory sensitivities, and anxiety in autistic and typically developing children. *Journal of Autism and Developmental Disorders*, *46*(6), 1962–1973.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*(4), 313–324.

Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, *54*(2), 638.

Nicolaidis, C., Kripke, C. C., & Raymaker, D. (2014). Primary care for adults on the autism spectrum. *Medical Clinics of North America*, *98*(5), 1169–1191.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.

Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1–18). New York: Plenum.

Nowell, K. P., Brewton, C. M., Allain, E., & Mire, S. S. (2015). The influence of demographic factors on the identification of autism spectrum disorder: a review and call for research. *Review Journal of Autism and Developmental Disorders*, *2*(3), 300–309.

O'Reilly, C., Lewis, J. D., Elsabbagh, M., O'Reilly, C., Lewis, J. D., Elsabbagh, M., … Elsabbagh, M. (2017). Is functional brain connectivity atypical in autism? a systematic review of eeg and meg studies. *PLoS ONE*, *12*(5), 1–28.

Olde Dubbelink, L. M. E., & Geurts, H. M. (2017). Planning skills in autism spectrum disorder across the lifespan: a meta-analysis and meta-regression. *Journal of Autism and Developmental Disorders*, *47*(4), 1148–1165.

Oliver, M. (2013). The social model of disability: thirty years on. *Disability and Society*, *28*(7), 1024–1026.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*(10), 1021–1034.

Ozonoff, S. J. (1995). Executive functions in autism. In *Learning and Cognition in Autism* (pp. 199–219).

Ozonoff, S. J., Cook, I., Coon, H., Dawson, G., Joseph, R. M., Klin, A., … Wrathall, D.

(2004). Performance on cambridge neuropsychological test automated battery subtests sensitive to frontal lobe function in people with autistic disorder: evidence from the collaborative programs of excellence in autism network. *Journal of Autism and Developmental Disorders*, *34*(2), 139–150.

Ozonoff, S. J., & Jensen, J. (1999). Brief report: specific executive function profiles in three neurodevelopmental disorders. *Journal of Autism and Developmental Disorders*, *29*(2), 171–177.

Ozonoff, S. J., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology*, *6*(1994), 415–431.

Ozonoff, S. J., Rogers, S. J., & Pennington, B. F. (1991). Asperger's syndrome: evidence of an empirical distinction from high-functioning autism. *Journal of Child Psychology and Psychiatry*, *32*(7), 1107–1122.

Ozonoff, S. J., & Strayer, D. L. (1997). Inhibitory function in nonretarded children with autism. *Journal of Autism and Developmental Disorders*, *27*(1), 59–77.

Ozonoff, S. J., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders*, *31*(3), 257–263.

Ozonoff, S. J., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive function abilities in autism and tourette syndrome: an information processing approach. *Journal of Child Psychology and Psychiatry*, *35*(6), 1015–1032.

Ozsivadjian, A., Knott, F., & Magiati, I. (2012). Parent and child perspectives on the nature of anxiety in children and young people with autism spectrum disorders: a focus group study. *Autism*, *16*(2), 107–121.

Panerai, S., Tasca, D., Ferri, R., Genitori D'Arrigo, V., & Elia, M. (2014). Executive functions and adaptive behaviour in autism spectrum disorders with and without intellectual disability. *Psychiatry Journal*, *2014*, 941809.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, *34*(5), 1189–208.

Peirce, J. W. (2007). PsychoPy - psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13.

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: clues to developmental primacy. *Developmental Psychology*, *43*(4), 974–990.

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental Psychology*, *46*(2), 530–544.

Pellicano, E. (2011). Psychological models of autism: an overview. In I. Roth & P. Rezaie (Eds.), *Researching the Autism Spectrum: Contemporary Perspectives* (pp. 219–249). Cambridge: Cambridge University Press.

Pellicano, E. (2012). The development of executive function in autism. *Autism Research and Treatment*, *2012*, 1–8.

Pellicano, E. (2013). Testing the predictive power of cognitive atypicalities in autistic children: evidence from a 3-year follow-up study. *Autism Research*, *6*(4), 258–267.

Pellicano, E., & Burr, D. (2012). When the world becomes "too real": a bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510.

Pellicano, E., Cribb, S., & Kenny, L. (2018). "It's changed a lot since then": patterns of continuity and change in the psychosocial outcomes of young autistic adults followed from childhood. *Under Submission.*

Pellicano, E., Kenny, L., Brede, J., Klaric, E., Lichwa, H., & McMillin, R. (2017). Executive function predicts school readiness in autistic and typical preschool children. *Cognitive Development*, *43*, 1–13.

Pellicano, E., Maybery, M. T., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder:" weak" central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology*, *18*(1), 77.

Pellicano, E., Ne'eman, A., & Stears, M. (2011). Engaging, not excluding: a response to walsh et al. *Nature Reviews Neuroscience*, *12*(12), 1.

Penney, S. (2013). Qualitative investigation of school-related issues affecting individuals diagnosed with autism spectrum disorder and co-occurring anxiety and/or depression. *Autism Insights*, *5*(1), 75–91.

Pennington, B. F., Rogers, S. J., Bennetto, L., Griffith, E. M., Reed, D. T., & Shyu, V. (1997). Validity tests of the executive dysfunction hypothesis of autism. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 143–178). Oxford: Oxford University Press.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, *20*(3), 249–262.

Philip, R. C. M., Dauvermann, M. R., Whalley, H. C., Baynham, K., Lawrie, S. M., & Stanfield, A. C. (2012). A systematic review and meta-analysis of the fmri investigation of autism spectrum disorders. *Neuroscience and Biobehavioral Reviews*, *36*(2), 901–942.

Piaget, J. (1954). *The construction of reality in the child.* New York, NY: Basic Books.

Piaget, J. (1978). *The grasp of consciousness. Action and concept in the young child*. London: Psychology Press.

Piening, S., van Dijk, A., Mirzaeva, N., Huizinga, P., Vieira Rocha, M., P., D. S., & van Balkom, I. D. (2018). Is eye movement desensitization and reprocessing effective in adults with an autism spectrum disorder and comorbid post traumatic stress disorder? In *A poster presentation at The Annual Meeting of the International Society for Autism Research (INSAR), May, 2018*. Rotterdam, NL.

Pierrot-Deseilligny, C., Müri, R. M., Ploner, C. J., Gaymard, B., Demeret, S., & Rivaud-Pechoux, S. (2003). Decisional role of the dorsolateral prefrontal cortex in ocular motor behaviour. *Brain*, *126*(6), 1460–1473.

Planche, P., & Lemonnier, E. (2012). Children with high-functioning autism and asperger's syndrome: can we differentiate their cognitive profiles? *Research in Autism Spectrum Disorders*, *6*(2), 939–948.

Poldrack, R. A. (2010). Mapping mental function to brain structure: how can cognitive neuroimaging succeed? *Perspectives on Psychological Science*, *5*(6), 753–761.

Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual Review of Psychology*, *67*(1), 587–612.

Pope, & Mays. (1995). Qualitative research: reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ*, *311*, 42–45.

Possin, K. L., Kang, G. A., Guo, C., Fine, E. M., Trujillo, A. J., Racine, C. A., … Kramer, J. H. (2013). Rivastigmine is associated with restoration of left frontal brain activity in parkinson's disease. *Movement Disorders*, *28*(10), 1384–1390.

Possin, K. L., LaMarre, A. K., Wood, K. A., Mungas, D. M., & Kramer, J. H. (2014). Ecological validity and neuroanatomical correlates of the nih examiner executive composite score. *Journal of the International Neuropsychological Society*, *20*(1), 20–28.

Potvin, M.-C., Snider, L., Prelock, P. A., Wood-Dauphinee, S., & Kehayia, E. (2015). Health-related quality of life in children with high-functioning autism. *Autism*, *19*(1), 14–19.

Preece, D., & Jordan, R. (2010). Obtaining the views of children and young people with autism spectrum disorders about their experience of daily life and social care support. *British Journal of Learning Disabilities*, *38*(1), 10–20.

Premack, D., & Woodruff, G. (1978). Doe the chimpanzee have a theory of mind? *The Behavioural and Brain Sciences*, *4*(1978), 515–526.

Prior, M., & Hoffmann, W. (1990). Brief report: neuropsychological testing of autistic children through an exploration with frontal lobe tests. *Journal of Autism and Developmental Disorders*, *20*(4), 581–590.

Pugliese, C. E., Anthony, L. G., Strang, J. F., Dudley, K., Wallace, G. L., & Kenworthy, L. E. (2015). Increasing adaptive behavior skill deficits from childhood to adolescence in autism spectrum disorder: role of executive function. *Journal of Autism and Developmental Disorders*, *45*(6), 1579–1587.

Pugliese, C. E., Anthony, L. G., Strang, J. F., Dudley, K., Wallace, G. L., Naiman, D. Q., & Kenworthy, L. E. (2016). Longitudinal examination of adaptive behavior in autism spectrum disorders: influence of executive function. *Journal of Autism and Developmental Disorders*, *46*(2), 467–477.

Pullinger, J. M. (2017). *Developing and piloting an ecologically valid measure of executive function for children with autism: a function-led approach*. UCL (University College London).

R Core Team. (2017). R: a language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Rabbitt, P. (2004). Do "frontal tests" measure executive function? issues of assessment and evidence from fluency tests. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 193–215). London: Routledge.

Rahko, J. S., Vuontela, V. A., Carlson, S. S., Nikkinen, J., Hurtig, T. M., Kuusikko-Gauffin, S., … Kiviniemi, V. J. (2016). Attention and working memory in adolescents with autism spectrum disorder: a functional mri study. *Child Psychiatry and Human Development*, *47*(3), 503–517.

Rajendran, G., Law, A. S., Logie, R. H., Van Der Meulen, M., Fraser, D., & Corley, M. (2011). Investigating multitasking in high-functioning adolescents with autism spectrum disorders using the virtual errands task. *Journal of Autism and Developmental Disorders*, *41*(11), 1445–1454.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review*, *27*(2), 224–260.

Rajendran, G., Mitchell, P., & Rickards, H. (2005). How do individuals with asperger syndrome respond to nonliteral language and inappropriate requests in computer-mediated communication? *Journal of Autism and Developmental Disorders*, *35*(4), 429–443.

Rand, D., Rukan, S. B.-A. A., Weiss, P. L., & Katz, N. (2009). Validation of the virtual met as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, *19*(4), 583–602.

Rauch, A., Lückenkemper, M., & Cieza, A. (2012). Introduction to the international classification of functioning, disability and health. In *ICF core sets: manual for clinical practice.* (pp. 4–13). Göttingen: Hogrefe Publishing.

Raver, C. C., Blair, C., & Willoughby, M. (2012). Poverty as a predictor of 4-year-olds ’executive function: new perspectives on models of differential susceptibility. *Developmental Psychology*, *49*(2), 292.

Raymaekers, R., Antrop, I., Van Der Meere, J., Wiersema, J. R., & Roeyers, H. (2007). HFA and adhd: a direct comparison on state regulation and response inhibition. *Journal of Clinical and Experimental Neuropsychology*, *29*(4), 418–427.

Raymaekers, R., Van Der Meere, J., & Roeyers, H. (2006). Response inhibition and immediate arousal in children with high-functioning autism. *Child Neuropsychology*, *12*(4–5), 349–359.

Remington, A. M., Swettenham, J. G., Campbell, R., & Coleman, M. (2009). Selective attention and perceptual load in autism spectrum disorder. *Psychological Science*, *20*(11), 1388–1393.

Remington, A. M., Swettenham, J. G., & Lavie, N. (2012). Lightening the load: perceptual load impairs visual detection in typical adults but not in autism. *Journal of Abnormal Psychology*, *121*(2), 544–551.

Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2012). The ecological and construct validity of a newly developed measure of executive function: the virtual library task. *Journal of the International Neuropsychological Society*, *18*(3), 440–450.

Rice, C. E., Rosanoff, M., Dawson, G., Durkin, M. S., Croen, L. A., Singer, A., & Yeargin-Allsopp, M. (2012). Evaluating changes in the prevalence of the autism spectrum disorders (asds). *Public Health Rev*, *34*(2), 1–22.

Rizopoulos, D. (2006). Ltm: an r package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., & Rabbitt, P. (1998). A study of performance on tests from the cantab battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers : implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society*, *4*(5), 474–490.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (2010). Cambridge neuropsychological test automated battery (cantab): a factor analytic study of a large sample of normal elderly volunteers. *Dementia and Geriatric Cognitive Disorders*, *5*(5), 266–281.

Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F. G. E., & Ronald, A. (2013). Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(13), 5258–62.

Robinson, H., Calamia, M., Gläscher, J., Bruss, J., & Tranel, D. (2014). Neuroanatomical correlates of executive functions: a neuropsychological approach using the examiner battery. *Journal of the International Neuropsychological Society*, *20*(01), 52–63.

Robinson, S., Goddard, L., Dritschel, B., Wisley, M., & Howlin, P. (2009). Executive functions in children with autism spectrum disorders. *Brain and Cognition*, *71*(3), 362–368.

Ronemus, M., Iossifov, I., Levy, D., & Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. *Nature Reviews. Genetics*, *15*(2), 133–41.

Rosenberg, R. E., Law, J. K., Yenokyan, G., McGready, J., Kaufmann, W. E., & Law, P. A. (2009). Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Archives of Pediatrics & Adolescent Medicine*, *163*(10), 907–914.

Rosenthal, M., Wallace, G. L., Lawson, R. A., Wills, M. C., Yerys, B. E., Kenworthy, L. E., & Dixon, E. (2013). Impairments in real-world executive function increase from childhood to adolescence in autism spectrum disorders. *Neuropsychology*, *27*(1), 13–18.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*(5), 343–350.

RStudio Team. (2015). RStudio: integrated development for r. Boston, MA: RStudio, Inc.

Rubenstein, E., & Bishop-Fitzpatrick, L. (2018). A matter of time: the necessity of temporal language in research on health conditions that present with autism spectrum disorder. *Autism Research*, 1–6.

Ruble, L. A., & Dalrymple, N. J. (1996). An alternative view of outcome in autism. *Focus on Autism and Other Developmental Disabilities*, *11*(1), 3–14.

Rumsey, J. M., & Hamburger, S. D. (1988). Neuropsychological findings in high-functioning men with infantile autism, residual state. *Journal of Clinical and Experimental Neuropsychology*, *10*(2), 201–221.

Russell-Smith, S. N., Comerford, B. J. E., Maybery, M. T., & Whitehouse, A. J. O. (2014). Brief report: further evidence for a link between inner speech limitations and executive function in high-functioning children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(5), 1236–1243.

Russell, J. (1997). *Autism as an executive disorder*. Oxford University Press.

Russell, J. (2013). *Agency: Its role in mental development*. Psychology Press.

Russell, J., Hala, S., & Hill, E. (2003). The automated windows task: the performance of preschool children, children with autism, and children with moderate learning difficulties. *Cognitive Development*, *18*(1), 111–137.

Russell, J., Jarrold, C., & Henry, L. A. (1996). Working memory in children with autism and with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *37*(6), 673–686.

Russell, J., Jarrold, C., & Hood, B. (1999). Two intact executive capacities in children with autism: implications for the core executive dysfunctions in the disorder. *Journal of Autism and Developmental Disorders*, *29*(2), 103–112.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of*

Developmental Psychology, *9*(2), 331–349.

Rutherford, M. D., & Rogers, S. J. (2003). Cognitive underpinnings of pretend play in autism. *Journal of Autism and Developmental Disorders*, *33*(3), 289–302.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual.* Western Psychological Services.

Ruxton, G. D., & Neuha, M. (2010). When should we use one-tailed hypothesis testing ?, 114–117.

Sachse, M., Schlitt, S., Hainz, D., Ciaramidaro, A., Schirman, S., Walter, H., … Freitag, C. M. (2013). Executive and visuo-motor function in adolescents and adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *43*(5), 1222–1235.

Saggers, B., Hwang, Y.-S., & Mercer, K. L. (2011). Your voice counts: listening to the voice of high school students with autism spectrum disorder. *Australasian Journal of Special Education*, *35*(02), 173–190.

Salomone, E., Shephard, E., Milosavljevic, B., Johnson, M. H., Charman, T., & The BASIS Team. (2018). Adaptive behaviour and cognitive skills: stability and change from 7 months to 7 years in siblings at high familial risk of autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(9), 2901–2911.

Sanderson, C., & Allen, M. L. (2013). The specificity of inhibitory impairments in autism and their relation to adhd-type symptoms. *Journal of Autism and Developmental Disorders*, *43*(5), 1065–1079.

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA: The Journal of the American Medical Association*, *311*(17), 1770–7.

Schatz, J., Stancil, M., Katz, T., & Sanchez, C. E. (2014). EXAMINER executive function battery and neurologic morbidity in pediatric sickle cell disease. *Journal of the International Neuropsychological Society*, *20*(1), 29–40.

Schmidt, L., Kirchner, J., Strunz, S., Broźus, J., Ritter, K., Roepke, S., & Dziobek, I. (2015). Psychosocial functioning and life satisfaction in adults with autism spectrum disorder without intellectual impairment. *Journal of Clinical Psychology*, *71*(12), 1259–1268.

Schmidt, T., Duncan, S., Ehmer, O., Kipp, M., Loehr, D., Magnusson, M., … Sloetjes, H. (2008). An exchange format for multimodal annotations. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Retrieved from http://www.lrec-conf.org/proceedings/lrec2008/

Schmitz, N., Rubia, K., Daly, E. M., Smith, A., Williams, S. C. R., & Murphy, D. G. M. (2006). Neural correlates of executive function in autistic spectrum disorders. *Biological Psychiatry*, *59*(1), 7–16.

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, *129*(2), 410–417.

Schneider, M., Eliez, S., Birr, J., Menghetti, S., Debbané, M., & Van der Linden, M. (2016). Multitasking abilities in adolescents with 22q11.2 deletion syndrome: results from an experimental ecological paradigm. *American Journal on Intellectual and Developmental Disabilities*, *121*(2), 151–164.

Schreiber, J. E., Possin, K. L., Girard, J. M., & Rey-Casserly, C. (2014). Executive function in children with attention deficit/hyperactivity disorder: the nih examiner battery. *Journal of the International Neuropsychological Society*, *20*(1), 41–51.

Schurink, J., Hartman, E., Scherder, E. J. A., Houwen, S., & Visscher, C. (2012). Relationship between motor and executive functioning in school-age children with pervasive developmental disorder not otherwise specified. *Research in Autism Spectrum*

Disorders, *6*(2), 726–732.

Schutz, S. E. (1994). Exploring the benefits of a subjective approach in qualitative nursing research. *Journal of Advanced Nursing*, *20*(3), 412–417.

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: a replication attempt. *Royal Society Open Science*, *5*(5), 172273.

Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit theory of mind reasoning in autism spectrum disorders: the impact of experience. *Autism*, *19*(4), 459–468.

Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (Vol. 1, pp. 118–137). Thousand Oaks, CA, US: SAGE Publications, Inc.

Schwartz, M. F., Reed, E. S., Montgomery, M., Palmer, C., & Mayer, N. H. (1991). *The quantitative description of action disorganisation after brain damage: A case study*. *Cognitive Neuropsychology* (Vol. 8).

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.

Seltzer, M. M., Shattuck, P. T., Abbeduto, L., & Greenberg, J. S. (2004). Trajectory of development in adolescents and adults with autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(4), 234–247.

Semrud-Clikeman, M., Walkowiak, J., Wilkinson, A., & Butcher, B. (2010). Executive functioning in children with asperger syndrome, adhd-combined type, adhd-predominately inattentive type, and controls. *Journal of Autism and Developmental Disorders*, *40*(8), 1017–1027.

Senju, A., Southgate, V., White, S. J., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in asperger syndrome. *Science*, *325*(5942), 883–885.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London*, *298*, 199–209.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727–741.

Shattuck, P. T. (2006). The contribution of diagnostic substitution to the growing administrative prevalence of autism in us special education. *Pediatrics*, *117*(4), 1028–1037.

Shattuck, P. T., Orsmond, G. I., Wagner, M., & Cooper, B. P. (2011). Participation in social activities among adolescents with an autism spectrum disorder. *PLoS ONE*, *6*(11).

Shattuck, P. T., Seltzer, M. M., Greenberg, J. S., Orsmond, G. I., Bolt, D., Kring, S., … Lord, C. (2007). Change in autism symptoms and maladaptive behaviors in adolescents and adults with an autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*(9), 1735–1747.

Shipman, D. L., Sheldrick, R. C., & Perrin, E. C. (2011). Quality of life in adolescents with autism spectrum disorders: reliability and validity of self-reports. *Journal of Developmental & Behavioral Pediatrics*, *32*(2), 85–89.

Shoda, Y., & Mischel, W. (1990). Predicting adolescent cognative and self-regulatory competencies from preeschool delay of gratification: identfying diagnostic conditions. *Developmental Psychology*, *26*(6), 978–986.

Shukla, D. K., Keehn, B., Smylie, D. M., & Müller, R.-A. (2011). Microstructural abnormalities of short-distance white matter tracts in autism spectrum disorder. *Neuropsychologia*, *49*(5), 1378–1382.

Simonoff, E., Jones, C. R. G., Baird, G., Pickles, A., Happé, F. G. E., & Charman, T. (2013). The persistence and stability of psychiatric problems in adolescents with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *54*(2), 186–194.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of American Academy of Adolescent Psychiatry*, *47*(8), 921–929.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Review*, *17*(3), 103–186.

Sinclair, J. (1999). Why i dislike 'person-first' language. Retrieved October 30, 2014, from http://web.archive.org/ web/20090210190652/http://web.syr.edu/~jisincla/person_ first.htm (accessed

Singer, J. (1999). "Why can't you be normal for once in your life?" from a "problem with no name" to the emergence of a new category of difference. In J. Singer & S. French (Eds.), *Disability Discourse* (Vol. Disability, pp. 59–67). Buckingham: Open University Press.

Sinzig, J., Morsch, D., Bruning, N., Schmidt, M. H., & Lehmkuhl, G. (2008). Inhibition, flexibility, working memory and planning in autism spectrum disorders with and without comorbid adhd-symptoms. *Child and Adolescent Psychiatry and Mental Health*, *2*(1), 4.

Skuse, D. H. (2007). Rethinking the nature of genetic vulnerability to autistic spectrum disorders. *Trends in Genetics*, *23*(8), 387–395.

Skuse, D. H., Mandy, W., Steer, C., Miller, L. L., Goodman, R., Lawrence, K., … Golding, J. (2009). Social communication competence and functional adaptation in a general population of children: preliminary evidence for sex-by-verbal iq differential risk. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(2), 128–137.

Smith, R. S., & Sharp, J. (2013). Fascination and isolation: a grounded theory exploration of unusual sensory experiences in adults with asperger syndrome. *Journal of Autism and Developmental Disorders*, *43*(4), 891–910.

Smith, T., Noble, M., Noble, S., Wright, G., McLennan, D., & Plunkett, E. (2015). *The English Indices of Deprivation 2015: Research report*. London: Department for Communities and Local Government. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464597/English_Indices_of_Deprivation_2015_-_Research_Report.pdf

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, *6*, 328.

Soke, G. N., Maenner, · M J, Christensen, · D, Kurzius-Spencer, · M, & Schieve, L. A. (2018). Prevalence of co-occurring medical and behavioral conditions/symptoms among 4-and 8-year-old children with autism spectrum disorder in selected areas of the united states in 2010. *Journal of Autism and Developmental Disorders*, *48*, 2663–2676.

Solomon, M., Ozonoff, S. J., Cummings, N., & Carter, C. S. (2008). Cognitive control in autism spectrum disorders. *International Journal of Developmental Neuroscience*, *26*(2), 239–247.

South, M., Larson, M. J., Krauskopf, E., & Clawson, A. (2010). Error processing in high-functioning autism spectrum disorders. *Biological Psychology*, *85*(2), 242–251.

South, M., Ozonoff, S. J., & Mcmahon, W. M. (2007). The relationship between executive functioning, central coherence, and repetitive behaviors in the high-functioning autism spectrum. *Autism*, *11*(5), 437–451.

Sparrow, M. (2016). Autistic inertia: an overview. Retrieved from http://unstrangemind.com/autistic-inertia-an-overview/

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales:(Vineland II), survey interview form/caregiver rating form*. Livonia, MN: Pearson Assessments.

Spek, A. A., Schatorjé, T., Scholte, E., & van Berckelaer-Onnes, I. (2009). Verbal fluency in adults with high functioning autism or asperger syndrome. *Neuropsychologia*, *47*(3), 652–656.

Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, *36*(5), 545–566.

Spiers, H. J., & Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *NeuroImage*, *31*(4), 1826–1840.

Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: a case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, *21*(4), 327–337.

Stahl, L., & Pry, R. (2002). Joint attention and set-shifting in young children with autism. *Autism*, *6*(4), 383–396.

Steel, J. G., Gorman, R., & Flexman, J. E. (1984). Neuropsychiatric testing in an autistic mathematical idiot-savant: evidence for nonverbal abstract capacity. *Journal of the American Academy of Child Psychiatry*, *23*(6), 704–7.

Steele, S. D., Minshew, N. J., Luna, B., & Sweeney, J. A. (2007). Spatial working memory deficits in autism. *Journal of Autism and Developmental Disorders*, *37*(4), 605–612.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (ema) in behavorial medicine. *Annals of Behavioral Medicine*.

Strang, J. F., Anthony, L. G., Yerys, B. E., Hardy, K. K., Wallace, G. L., Armour, A. C., … Kenworthy, L. E. (2017). The flexibility scale: development and preliminary validation of a cognitive flexibility measure in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *47*(8), 2502–2518.

Strang, J. F., Kenworthy, L. E., Daniolos, P., Case, L., Wills, M. C., Martin, A., & Wallace, G. L. (2012). Depression and anxiety symptoms in children and adolescents with autism spectrum disorders without intellectual disability. *Research in Autism Spectrum Disorders*, *6*(1), 406–412.

Strayer, D. L., Medeiros-Ward, N., & Watson, J. M. (2013). Gender invariance in multitasking: a comment on mäntylä (2013). *Psychological Science*, *24*(5), 809–810.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York, NY: Raven Press.

Sullivan, A. (2002). Inertia: from theory to praxis. Retrieved from https://web.archive.org/web/20160301144431/http://archive.autistics.org/library/inertia.html

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1988). Decision processes in perception. *Adolescent Gynecology*, *84*(1967), 1932–1935.

Szameitat, A. J., Hamaida, Y., Tulley, R. S., Saylik, R., & Otermans, P. C. J. (2015). “Women are better than men”: public beliefs on gender differences and other aspects in multitasking. *PLoS ONE*, *10*(10), 2–6.

Taylor, B., Jick, H., & MacLaughlin, D. (2013). Prevalence and incidence rates of autism in the uk: time trend from 2004-2010 in children aged 8 years. *BMJ Open*, *3*(10).

Testa, R., Bennett, P., & Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Archives of Clinical Neuropsychology*, *27*(2), 213–224.

The Centers for Disease Control and Prevention. (2007). Evaluation of a methodology for a collaborative multiple source surveillance network for autism spectrum disorders—autism and developmental disabilities monitoring network, 14 sites, united states, 2002. *Morbidity and Mortality Weekly Report. Surveillance Summaries/CDC*, *56*(1), 29–40.

The Centers for Disease Control and Prevention. (2009). Prevalence of autism spectrum disorders --- autism and developmental disabilities monitoring network, united states, 2006. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, *58*(SS10), 1–20.

The Centers for Disease Control and Prevention. (2012). Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, united states, 2008. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, *61*(3), 1–19.

The Centers for Disease Control and Prevention. (2014). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2010. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, *63*(2), 1–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24670961

The Centers for Disease Control and Prevention. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years-- autism and developmental disabilities monitoring network, 11 sites, united states, 2012. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, *65*(3), 1–23.

The Centers for Disease Control and Prevention. (2017). Checklist: signs and symptoms of attention-deficit/hyperactivity disorder (adhd). Retrieved from http://addresources.org/cdc-adhd-symptom-checklist-for-children/

Thomas, C. (2004). Developing the social relational in the social model of disability: a theoretical agenda. In C. Barnes & G. Mercer (Eds.), *Implementing the Social Model of*

Disability: Theory and Research (Vol. Leeds, pp. 32–47). Leeds, UK: Disability Press.

Thomas, M. S. C., Davis, R., Karmiloff-smith, A., Knowland, V. C. P., & Charman, T. (2016). The over-pruning hypothesis of autism. *Developmental Science*, *19*(2), 284–305.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(2), 131–143.

Towgood, K. J., Meuwese, J. D. I., Gilbert, S. J., Turner, M. S., & Burgess, P. W. (2009). Advantages of the multiple case series approach to the study of cognitive deficits in autism spectrum disorder. *Neuropsychologia*, *47*(13), 2981–2988.

Tsai, P., & Sahin, M. (2011). Mechanisms of neurocognitive dysfunction and therapeutic considerations in tuberous sclerosis complex. *Current Opinion in Neurology*, *24*(2), 106–13.

Tsuchiya, E., Oki, J., Yahara, N., & Fujieda, K. (2005). Computerized version of the wisconsin card sorting test in children with high-functioning autistic disorder or attention-deficit/hyperactivity disorder. *Brain and Development*, *27*(3 SPEC. ISS.), 233–236.

Tupper, D. E., & Cicerone, K. D. (1990). Introduction to the neuropsychology of everyday life. In *The neuropsychology of everyday life: Assessment and basic competencies* (pp. 3–18). Springer.

Unterrainer, J. M., Rauh, R., Rahm, B., Hardt, J., Kaller, C. P., Klein, C., … Biscaldi, M. (2016). Development of planning in children with high-functioning autism spectrum disorders and/or attention deficit/hyperactivity disorder. *Autism Research*, *9*(7), 739–751.

Urbain, C., Pang, E. W., & Taylor, M. J. (2015). Atypical spatiotemporal signatures of working memory brain processes in autism. *Translational Psychiatry*, *5*(8), e617-9.

van de Wettering, J. (2010). Are special interests another form of autistic inertia? Retrieved September 20, 2018, from http://web.archive.org/web/20130515103516/http://www.lettersfromaspergia.com/2010/11/are-special-interests-another-form-of.html

van den Bergh, S. F. W. M., Scheeren, A. M., Begeer, S., Koot, H. M., & Geurts, H. M. (2014). Age related differences of executive functioning problems in everyday life of children and adolescents in the autism spectrum. *Journal of Autism and Developmental Disorders*, *44*(8), 1959–1971.

Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W., & Wagemans, J. (2015). Global processing takes time: a meta-analysis on local–global visual processing in asd. *Psychological Bulletin*, *141*(3), 549–573.

Van Eylen, L., Boets, B., Steyaert, J., Wagemans, J., & Noens, I. (2015). Executive functioning in autism spectrum disorders: influence of task and sample characteristics and relation to symptom severity. *European Child and Adolescent Psychiatry*, *24*(11), 1399–1417.

Varni, J. W., Seid, M., & Kurtin, P. S. (2001). PedsQL™ 4.0: reliability and validity of the pediatric quality of life inventory ™ version 4.0 generic core scales in healthy and patient populations. *Medical Care*, *39*(8), 800–812.

Vaskinn, A., Sergi, M. J., & Green, M. F. (2009). The challenges of ecological validity in the measurement of social perception in schizophrenia. *The Journal of Nervous and Mental Disease*, *197*(9), 700–702.

Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-it: windows executable software for the stop-signal paradigm. *Behavior Research Methods*, *40*(2), 479–483.

Verté, S., Geurts, H. M., Roeyers, H., Oosterlaan, J., & Sergeant, J. A. (2006). Executive functioning in children with an autism spectrum disorder: can we differentiate within the spectrum? *Journal of Autism and Developmental Disorders*, *36*(3), 351–372.

Verté, S., Geurts, H. M., Roeyers, H., Oosterlaan, J., & Sergeant, J. a. (2005). Executive functioning in children with autism and tourette syndrome. *Development and Psychopathology*, *17*(2), 415–445.

Visser, E. M., Berger, H. J. C., Lantman-De Valk, H. M. J. V. S., Prins, J. B., & Teunisse, J.-P. (2015). Cognitive shifting and externalising problem behaviour in intellectual disability and autism spectrum disorder. *Journal of Intellectual Disability Research*, *59*(8), 755–766.

Visu-Petra, L., Cheie, L., Benga, O., & Miclea, M. (2012). The structure of executive functions in preschoolers: an investigation using the nepsy battery. *Procedia - Social and Behavioral Sciences*, *33*, 627–631.

Volkmar, F. R., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., & State, M. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(2), 237–257.

Vriezen, E. R., & Pigott, S. E. (2016). The relationship between parental report on the brief and performance-based measures of executive function in children with moderate to severe traumatic brain injury. *Child Neuropsychology*, *7049*(July), 296–303.

Wallace, G. L., Kenworthy, L. E., Pugliese, C. E., Popal, H. S., White, E. I., Brodsky, E., & Martin, A. (2016). Real-world executive functions in adults with autism spectrum disorder: profiles of impairment and associations with adaptive functioning and co-morbid anxiety and depression. *Journal of Autism and Developmental Disorders*, *46*(3).

Wang, Y., Zhang, Y., Liu, L., Cui, J., Wang, J., Shum, D. H. K., … Chan, R. C. K. (2017). A meta-analysis of working memory impairments in autism spectrum disorders. *Neuropsychology Review*, *27*(1), 46–61.

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: a conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, *29*(7), 1159–1177.

Wechsler, D. (1998). *Wechsler Memory Scale*. London: Psychological Corporation.

Wechsler, D. (2011). *WASI-II: Wechsler abbreviated scale of intelligence--*. Psychological Corporation.

Wechsler, D., Golombok, S., & Rust, J. (1992). *Manual for the Wechsler Intelligence Scale for Children—Third Edition UK*. London: The Psychological Corporation.

Werling, D. M., & Geschwind, D. H. (2013). Sex differences in autism spectrum disorders. *Current Opinion in Neurology*, *26*(2), 146–153.

White, S. J. (2013). The triple i hypothesis: taking another('s) perspective on executive dysfunction in autism. *Journal of Autism and Developmental Disorders*, *43*(1), 114–121.

White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on "open-ended" executive function tests in autism. *Autism Research*, *2*(3), 138–147.

White, S. W., Oswald, D., Ollendick, T. H., & Scahill, L. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Clinical Psychology Review*, *29*(3), 216–229.

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., … Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, *4*(1), 63–85.

WHO. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines* (Vol. 1). Geneva: World Health Organization.

WHO. (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: World Health Organization.

WHO. (2018). *International Classification of Diseases, 11th Revision (ICD-11)*. Geneva: World Health Organization.

Wigham, S., Rodgers, 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), 943–952.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological Psychiatry*, *57*, 1336–1346.

Williams, D., & Jarrold, C. (2013). Assessing planning and set-shifting abilities in autism: are experimenter-administered and computerised versions of tasks equivalent? *Autism Research*, *6*(6), 461–467.

Williams, D. L., Goldstein, G., Carpenter, P. A., & Minshew, N. J. (2005). Verbal and spatial working memory in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 747–756.

Williams, D. L., Goldstein, G., & Minshew, N. J. (2005). Impaired memory for faces and social scenes in autism: clinical implications of memory dysfunction. *Archives of Clinical Neuropsychology*, *20*(1), 1–15.

Williams, D. L., Goldstein, G., & Minshew, N. J. (2006). The profile of memory function in children with autism. *Neuropsychology*, *20*(1), 21–29.

Williams, D. L., Mazefsky, C. A., Walker, J. D., Minshew, N. J., & Goldstein, G. (2014). Associations between conceptual reasoning, problem solving, and adaptive ability in high-functioning autism. *Journal of Autism and Developmental Disorders*, *44*(11), 2908–2920.

Williams, E. I., Gleeson, K., & Jones, B. E. (2017). How pupils on the autism spectrum make sense of themselves in the context of their experiences in a mainstream school setting: a qualitative metasynthesis. *Autism*, 136236131772383.

Williams, J., Rickert, V., Hogan, J., Zolten, A. I., Satz, P., Elia, L. F. D., … Light, R. (1995). Children's color trails. *Archives of Clinical Neuropsychology*, *10*(3), 211–223.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome*. Flempton, Bury St Edmunds: Thames Valley Test Company.

Wing, L. (1981). Asperger's syndrome: a clinical account. *PSYCHOLOGICAL MEDICINE*, *11*(1), 115–129.

Wolff, J. J., Gerig, G., Lewis, J. D., Soda, T., Styner, M. A., Vachet, C., … Wolff, J. (2015). Altered corpus callosum morphology associated with autism over the first 2 years of life. *Brain*, *138*(7), 2046–2058.

Wood, R. L., & Liossi, C. (2006). The ecological validity of executive tests in a severely brain injured sample. *Archives of Clinical Neuropsychology*, *21*(5), 429–437.

Xiao, T., Xiao, Z., Ke, X., Hong, S., Yang, H., Su, Y., … Liu, Y. (2012). Response inhibition impairment in high functioning autism and attention deficit hyperactivity disorder: evidence from near-infrared spectroscopy data. *PloS ONE*, *7*(10), e46569.

Ye, Z., Li, Y., Fathi, A., Han, Y., Rozga, A., Abowd, G. D., & Rehg, J. M. (2012). Detecting eye contact using wearable eye-tracking glasses. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, (January), 699.

Yerys, B. E., Hepburn, S. L., Pennington, B. F., & Rogers, S. J. (2007). Executive function in preschoolers with autism: evidence consistent with a secondary deficit. *Journal of Autism and Developmental Disorders*, *37*(6), 1068–1079.

Yerys, B. E., Wallace, G. L., Harrison, B., Celano, M. J., Giedd, J. N., & Kenworthy, L. E.

(2009). Set-shifting in children with autism spectrum disorders: reversal shifting deficits on the intradimensional/extradimensional shift test correlate with repetitive behaviors. *Autism*, *13*(5), 523–538.

Zahavi, D. (2010). Complexities of self. *Autism*, *14*(5), 547–551.

Zalla, T., Plassiart, C., Pillon, B., Grafman, J., & Sirigu, A. (2001). Action planning in a virtual context after prefrontal cortex damage. *Neuropsychologia*, *39*(8), 759–770.

Zelazo, P. D. (2006). The dimensional change card sort (dccs): a method of assessing executive function in children. *Nature Protocols*, *1*(1), 297–301.

Zelazo, P. D., Jacques, S., Burack, J. A., & Frye, D. (2002). The relation between theory of mind and rule use: evidence from persons with autism-spectrum disorders. *Infant and Child Development*, *11*(2), 171–195.

Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell Handbook of Childhood Cognitive Development* (pp. 445–469). Oxford: Blackwell Publishing.

Zinke, K., Fries, E., Altgassen, M., Kirschbaum, C., Dettenborn, L., & Kliegel, M. (2010). Visuospatial short-term memory explains deficits in tower task planning in high-functioning children with autism spectrum disorder. *Child Neuropsychology*, *16*(3), 229–241.

Appendices

Appendix A. Scoring criteria for the MMEA

Goals achieved

						Only complete if activity administered pre instruction update		
		1	2	3	4	5	6	
Sandwiches indicate if completed prior to update [ ]								
Goal criteria	NA - no sandwich was made, no sandwich should have been made 0 - sandwich was not made and it should have been made 1 - an attempt was made to partially make a sandwich (e.g., bread was placed on plate but not buttered, bread was buttered but sides of bread were never put together) 2 - sandwich made according to the following criteria: includes either one slice of bread cut in half or folded over or two slices of bread put together as well as at least one of the following fillings: butter, (vegetarian) ham, tomato, cheese,							Prop. Achieved
	Goal achieved							
Descriptors	2 sides of bread used							
	At least 1 side was buttered							
	(Vegetarian) ham included							
	Cheese included							
	Tomato included							
	Ketchup included							
	Sandwich cut in half/quarter							
	Placed on plate							
	Napkin on or next to plate							
	Cutlery laid out							
Quality	NA – Sandwich not made 1 – Sandwich was hastily assembled/presented poorly e.g., ingredients missing, butter not evenly spread, holes in bread from buttering, ingredients not evenly distributed 2 – sandwich made to a reasonable standard - e.g., contains correct ingredients but may be a little sloppy in appearance or rushed in preparation 3 – sandwich made to an excellent standard							Mean quality rating
	Quality rating							
Folders indicate if completed prior to update [ ]								
Goal criteria	NA - no folder was made, no folder should have been made 0 - folder not made and it should have been made 1 - an attempt was made to make a folder, e.g., one plastic folder was laid out with a single (un)stapled sheet or multiple unstapled sheets from the chapter were inserted/placed on a folder 2 - folder made according to the following criteria: A bundle is put together that includes a plastic folder and two or more stapled history chapter pages, pages may be placed on or inside folder							Prop. Achieved
	Goal achieved							
Descriptors	Plastic folder used							
	sheet(s) put inside folder							
	Included pen							
	Included notepad							
	Pages stapled							
	No page errors							
Quality	NA – folder not made 1 – an attempt made to compile folder but it is either presented poorly or participants used an inefficient strategy to compile. For example, participants may have inserted page(s) into the folder before stapling together or may not have checked that multiple pages were needed, the end result is an incomplete set of pages or pages containing many errors 2 – an attempt made to compile folder but participants used an inefficient strategy or it folder contains some errors. The end result is a complete set of pages but the strategy employed was inefficient 3 – Folders are made to an excellent standard and an efficient strategy is used, pages are in correct order, stapled and either on or in plastic folder and little time was wasted							Mean quality rating

Participant ID: __________ Coder initials: _________ Date coded: __________

	Quality rating							
Cups of tea indicate if completed prior to update [ ]								
Goal criteria	NA - no tea made, no tea should have been made 0 - No tea made, tea should have been made 1 - an attempt was made to make tea but it did not fulfill criteria below, e.g., cup was half-full, water had never been boiled, water was poured in before final 5 minutes 2 - cup of tea made according to the following criteria: kettle was switched on at some point during the task, water was poured within final 5 minutes of time limit, teabag is placed in cup, water poured in, tea is made with at least 2 minutes left on the clock, as instructed. Teabag may be left in or taken out. The order the teabag and water added are unimportant							Prop. Achieved
	Goal achieved							
Descriptors	One teabag was used							
	Cup was filled to the top							
	Kettle had been boiled at some point during task							
	Tea made with at least two minutes on the clock							
	Teabag was removed							
Quality	1 – tea is made but to a low standard, e.g., more than one of the following issues apply: the kettle was not boiled within the final 5 minutes of the task, the cup was only half filled, the tea was not stirred and teabag was not removed 2 – tea is made but not quite perfect but not as bad as for a score of 1, e.g., only one of the following issues apply: the kettle was not boiled within the final 5 minutes of the task, the cup was only half filled, the tea was not stirred and teabag was not removed 3 – Tea is made to an excellent standard							Mean quality rating
	Quality rating							
Table set								
Goal criteria	NA - no items were laid out, no items should have been laid out 0 - items are simply left in the place they were prepared, no attempt is made to lay them out intentionally for study session 1 - If some (but not all) of the folders, sandwiches and teas are placed together in bunches a point is given for each that is placed - showing intent to prepare for the study session OR a place is laid for an invidual attendee that contains some but not all of the following: tea, sandwich, folder 2 - Evidence that either all teas are placed together, all folders are placed together and all sandwiches are placed together							Prop. Achieved
	Goal achieved							
Descriptors	Placement of sandwich							
	Placement of folder							
	Placement of tea							
Quality	NA - place not set 1 – Some items are placed either on the side table or central table in an attempt to prepare for the study session, may not include all needed items or presentation of study area may be poor 2 – Table is set but is slightly messy or setting process was somewhat inefficient 3 – Table is set very neatly, all needed items are included. This may involve setting places with a sandwich, tea or folder for each person or placing all teas together, all folders together and all sandwiches together							Mean quality rating
	Quality rating							
	Narrative:							

Participant ID: __________ Coder initial: _________ Date coded: __________

Executive Errors and Omissions

	Count (if applicable)	Code		
Tea				
Kettle was boiled extra times				
Water not sufficiently hot (no/early boil)				
Teabags put in extra cups				
Water poured in extra cups				
Too few teas made (self omit not included)				
Tea poured back into kettle				
Tea not made within final 5 minutes of task				
Self not included before update				
Self not included after update				
Folders				
Folder not prepared as first activity				
Extra folders were made				
Page sequence errors				
Sheets put in folder before stapling				
too few folders made (self omit not included)				
Spent time reading history chapter				
Self not included before update				
Self not included after update				
Sandwiches				
Extra sandwiches made				
Too few sandwiches made (self omit not included)				
A sweet sandwich was made				
Bread for extra sandwiches was laid out/cut				
Self not included before update				
Self not included after update				
Table set				
Distractors laid on table				
Extra plates laid out				
Extra cups laid out				
Overall				
Finished more than a minute early				
Questions asked outside allotted time				
Ran out of time				
Omissions				
Tea omitted completely				
Folders omitted completely				
Sandwiches omitted completely				
Table set omitted completely				