Spotting Lies and Reading Minds:

Development of Mentalizing and Deception in

Autistic and Non-Autistic Individuals

Ishita Chowdhury

Thesis submitted to UCL for the degree of

Doctor of Philosophy

March 2023

Institute of Cognitive Neuroscience

Division of Psychology and Language Sciences

University College London

DECLARATION

I, Ishita Chowdhury, 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.

Signed: Ishita Chowdhury

Date: 27 March 2023

ABSTRACT

Deception is ever-present in day-to-day life. One cognitive process underlying deception, which has been observed to evolve throughout development, is mentalizing i.e. the ability to attribute mental-states to others. Autistic individuals have been found to struggle with mentalizing even in adulthood, so it is possible that they show difficulties in detecting deception as well. The main aims of this PhD were to investigate how mentalizing and deception develop in autistic and non-autistic individuals from pre-adolescence to early adulthood, and to investigate other factors that may affect deception judgement, specifically intergroup bias.

In my first study, I collected deception stimuli for two novel deception detection tasks, and investigated if mentalizing ability and autistic traits in a non-autistic sample were related to how successful one is at deceiving. I found that, contrary to expectations, deception production success did not correlate with either mentalizing or autistic traits.

For my second study, I tested 11-30 years old autistic and non-autistic participants, using a well-established detection paradigm as well as two novel deception detection tasks, and found that autistic individuals were weaker at detecting deception than non-autistic individuals. While both mentalizing and deception detection abilities improved with age in non-autistic individuals, neither were affected by age in autistic individuals. Furthermore, deception detection was found to predict peer-victimization, and through peer-victimization effect psychological distress.

For my final study, I investigated neurotype-based intergroup bias in the context of deception and found that, instead of better deception detection for the in-group (vs out-group) that was

expected, both autistic and non-autistic adults were better at detecting deception from other autistic adults.

I discuss the theoretical implications of these finding in terms of our understanding of the cognitive underpinnings of deception, the implications this has for autistic individuals' quality of life, and future avenues for deception and autism research.

Impact Statement

Deception is pervasive in our daily lives and while large-scale deception like fraud cost businesses and individuals £137 billion every year (Gee & Button, 2021), interpersonal deception on a more day-to-day basis also has dire consequences, especially for people who already face difficulties in social communication – such as autistic individuals. About 1% of the population is autistic, and autism’s economic cost is assessed at approximately £32 billion per year in the UK (Buescher et al., 2014). Autistic individuals report being bullied more than their non-autistic counterparts (Sterzing et al., 2012). As bullies often use deception as a means to peer-victimize others (Maboyana & Sekaja, 2015), weak deception detection skills could compound the issues that autistic individuals already face. This thesis aimed to investigate the development of deception in autistic and non-autistic individuals, and if mentalizing – i.e. the ability to attribute mental states – and intergroup bias affected deception detection as well. The findings of this project could have impact in several areas both inside and outside of academia.

Academic impact

Scholarship: Findings of this thesis added to the limited body of literature on development of mentalizing beyond childhood and how mentalizing relates to deception. This is also the first study to investigate how deception detection develops with age and how intergroup bias affects deception detection.

Methodological impact: Recognizing the shortcomings of past deception paradigms, two novel deception detection tasks were created that have potential to be used in future research. The findings highlighted the need for continued development of newer tasks to measure mentalizing. It was also showcased how deception detection tasks, on which autistic and non-autistic participants perform similarly to how they perform on traditional mentalizing tasks, have potential to be developed into measures of mentalizing ability.

Non-academic impact

Health and wellbeing: The findings indicated that deception detection likely contributes to vulnerability to peer-victimization and that this is especially important in autistic adults, since autistic individuals experience an increase in bullying victimization, as they grow older. This study was the first ever to demonstrate that deception detection also indirectly affects psychological distress and mental health through peer-victimization. These findings can be used to develop deception detection based anti-bullying interventions for those who are vulnerable – such as adolescents and autistic individuals.

Public discourse and quality of life: In order to improve autistic individuals' quality of life, the negative attitude toward autistic individuals found in this project underscores a need for interventions focussing on changing non-autistic perception of autism instead of merely focussing on changing autistic behaviour. This can be undertaken via increasing public awareness, possibly through campaigns, and knowledge exchange.

Influencing policy: The finding that autistic individuals are perceived to be more deceptive has far-reaching consequences within the criminal justice system. This could lead to autistic witnesses being dismissed and autistic suspects being judged wrongly and more harshly. These finding could be used to push for sensitivity training of individuals within the criminal justice system who are likely to encounter autistic individuals, such as police officers and first respondents.

ACKNOWLEDGEMENTS

It takes a village…

While my PhD was a solo endeavour, it was not a journey I undertook alone. So many people who have been with me from the beginning and so many others I have met on the way, have guided me, built me up, and made it possible for me to achieve this milestone today. Therefore, it is imperative that I try to put into words the gratitude I feel towards everyone who has contributed to this PhD in some or the other way.

First and foremost, to my wonderful supervisor, Dr Sarah White – I could not have hoped to find a better guide, a better mentor. Thank you so, so much for taking me under your wing and teaching me to think for myself. You recognized my potential before I did and opened doors for me that I could have only dreamt of. For that and so much more, I will forever be grateful. I consider myself incredibly lucky to have been at the receiving end of your wisdom, your expertise, your knowledge, but also your humour and the countless life lessons! Thank you for making me a part of your research family, the DevDivLab.

To my secondary supervisor Prof Sarah Jayne Blakemore for your guidance on the dos and don'ts of developmental research and Prof Antonia Hamilton for your generosity for offering alternative testing materials when the pandemic made it impossible for me to record my own. My thesis would have lacked some breadth and depth without both your help, thank you.

To the Masters students – Tikal Catena, Marina Geer, Imogen Krell, Catherine Hill, and Caitlin McDonald – who have slogged through the whole process with me. To all the participants who took part in my study, to all the schools and charities who were so kind to help us recruit even in the middle of a pandemic. To everyone else at DevDivLab who has always lent a helping hand. This project would not have been possible in its present form without you.

Now on to the friends who kept me sane behind the scene…

To my work wife, Kat Ellis, for being my sounding board, for always being ready for my rants and for pumping me with confidence with a well-placed "you got this" and picking me up with a "C'est la vie". To Beth Jones, for helping me take breaks by enticing me with Pret coffees. To my flatmate and family-away-from-home, Banika Ahuja, for reminding me to take it easy, eat, sleep, and take care of myself. The three of you have made being away from home, through this gruelling process, a lot easier – I cannot tell you how much I appreciate all your support.

To Ankita Marwaha and Vaanya Kaushik, two of my oldest friends, who have seen me through thick and thin and who have been silent cheerleaders for everything I do. Thank you for being there with me in spirit when you could not be there in person.

Last but not the least, to my family…

To my partner, Sidhant Verma for being an absolute rock and offering an ear and a shoulder on days when the going got tough, even from thousands of miles away. For putting up with my temper and my tears, with the limited communication these last six months as I kept mad hours. For the unconditional joy that you show in my successes. Your support, understanding, and patience are unmatched.

To my sister, Alekha Chowdhury, for sending me a million TikToks to help me unclench at the end of the day, for providing me reprieve from being an adult as we fangirled over Marvel and Disney. For always responding with a 'yes' to my 'you there?' text message, even at 4 in the morning.

To my dad, Prasun Chowdhury, for your unwavering faith in me. For your tough love and enthusiastic encouragement and everything in between. For inspiring me to work hard. To my mum, Mithu Chowdhury, for your unqualified support and care. For simply being there for me, whenever and however I needed you.

I literally do not have the words to tell you all how grateful I am for you. Thank you, for making me who I am today. I love you.

Finally…

To God, for never letting me lose sight of my goal.

Table Of Contents

List of Tables

Chapter 5

List of Figures

Chapter 2

Chapter 4

Chapter 5

Chapter 1. Introduction

It is a truth universally acknowledged that… people lie.[1]

We lie and experience being lied to on a daily basis. Your aunt gifts you an ugly China bowl and you enthuse about how much you like it (when you actually do not) just to spare her feelings. The family pet dies and you tell your children that it has gone to live on a farm far away to avoid upsetting them. White lies that are told with no ill intention, mostly to spare someone's feelings or maintain social equilibrium, are probably the most common forms of deception we encounter every day. Not all lies are harmless, though. It is worryingly common to receive spam emails telling us we have won a lottery or that some foreign prince has left us seven-figure inheritance. Similarly commonplace are constant warnings from banks about phishing attacks where fraudsters pose as trusted sources to scam unsuspecting people. Falling for these deceptions can have far-reaching consequences.

Deception can be a social navigation tool but deception can also be malicious. Either way, it cannot be denied that lies and deception are pervasive in everyday social interaction. They range from white lies told for prosocial and altruistic purposes as demonstrated above, to jokes used to assist in smoother communication, to feints in sports in order to throw off competitors, to antisocial lies told to hide true motives and deceive others for selfish gain such as forgeries and frauds (Abe, 2011). As such, considerable research has taken place trying to understand the psychological mechanism behind deception – how it is produced, how it is detected, and what the individual differences that help or hinder these processes are.

One cognitive concept underlying deception that has been identified is mentalizing, which is the ability to understand another person's mental state such as their beliefs and desires. Deception production requires an evaluation of what the other person may or may not know in order to then instil a false belief in them; deception detection requires the recognition that

[1]The first half of this statement is taken from opening line of the classic, Pride and Prejudice by Jane Austen. The second half is a generalisation based on DePaulo (2010)

someone intends to instil a false mental state in us (Talwar & Lee, 2008). For this reason, it can be premised that changes in mentalizing ability would lead to changes in deception abilities – both production and detection.

Not everybody has the same ability to understand other people's mental states. Mentalizing is thought to be weak in autism (Senju et al., 2009), therefore, it is possible that autistic individuals[2] also have difficulties in deception production and detection. Furthermore, mentalizing ability has been seen to change with age throughout adolescence (Blakemore, 2008), therefore it is likely that deception abilities change over time as well.

Therefore, my PhD project, as a whole, primarily aims to investigate how the interaction between mentalizing and deception develops from late childhood through adolescence to early adulthood, in autistic individuals and their non-autistic peers. It looks at how a weak deception detection ability can affect individuals and explores alternative explanations to why autistic people might struggle with deception detection.

In the first part of this chapter, I begin by exploring what deception is, how prevalent it is and how it is studied. I briefly touch upon the behavioural characteristics and components of deception and then expound the cognitive underpinnings of deception. The second part of this chapter focuses on mentalizing, also summarising our current understanding of how mentalizing and deception develop. In the third part, I introduce autism and outline our current understanding of autism, along with what we know about mentalizing and deception in autism. The last part of this chapter briefly brings up the implications of compromised deception detection. Finally, I conclude by outlining what the rest of this thesis is going to cover.

[2]Language statement: Throughout this thesis, I will be using identity-first language, as has been found to be preferred by a majority of the autistic community (Bottema-Beutel et al., 2021). Any mention of autistic participants in this project is in reference to autistic individuals who do not have co-occurring intellectual disabilities and are of average to high IQ.

1.1. Deception

1.1.1. What is deception?

From a psychological perspective, deception can be explained as a process in which the deceiver, in order to avoid some sort of loss or gain some sort of benefit, takes a fact that they know to be false and intentionally tries to make another individual believe it to be true (Abe, 2009). The variety of situations that deception is present in and the various different motives behind them demonstrate how pervasive and heterogeneous deception and lying actually are in daily life. Therefore, deception has been a popular area of research for psychologists.

The most commonly accepted definition of deception in psychological research is that deception is a bid to *purposely mislead another*. This definition eliminates instances where people lie by mistake, such as when they misremember details of an event, so the *intention* to deceive is missing; or instances where people self-deceive, such as when people are unaware of holding more than one belief, so they are not deceiving *another*. This definition of deception also eliminates instances where one may strategically lie with the intention to be caught, such as sarcasm; therefore, the intention to *mislead* is missing (Zuckerman et al., 1981). To add to this explanation, deception is “defined solely from the perspective of the deceiver and not from the factuality of the statement” (Vrij, 2008, p. 14). Therefore, a person delivering a truth while pretending that it is a lie is deceptive, while someone stating an incorrect fact that they believe to be true is not deceptive.

Deception also does not necessarily need to be verbal (Vrij, 2008): a child who is supposed to be studying but hides her mobile phone when she notices her parent enter the room is being deceptive without using words. It could be argued that deception could be verbal or non-verbal but lying is purely verbal, and some researchers like Bok (1978) insist that there is a difference between deceiving and lying. However, other more recent scholars in deception research, such as DePaulo et al. (2003), treat the two terms in their research as interchangeable, and that will be the norm in this project as well. Furthermore, in this project like in most deception literature (e.g.: C. F. Bond & DePaulo, 2008), the individual who is potentially producing deceptive behaviour will be referred to as a *sender* instead of 'liar' or 'deceiver' as they are as likely to send truthful signals or messages as they are to send deceptive signal or messages. Similarly, the person who is judging the sender's veracity will be referred to as a *receiver*.

1.1.2. Deception in research and real life

As demonstrated earlier, deception is pervasive in our daily lives. However, deception can go beyond just one-to-one instances of lying, and can have serious consequences. It is estimated that businesses and individuals collectively lose approximately £137 billion every year to fraud (Gee & Button, 2021). Serious instances of deception are often linked with serious crimes and researching deception in the context of the justice system has been deemed extremely important (Hartwig & Bond Jr., 2014). In fact, a lot of deception research in the last few decades has been fuelled by and is related to forensic psychology and the criminal justice system. It ranges from pinning down mechanisms of lie detection in legal scenarios (e.g.: Meissner & Kassin, 2002), to accuracy of law enforcement officers in deception-detection (e.g.: Mann et al., 2004), to improving interrogation techniques (Vrij et al., 2017), to training law enforcement to better detect lies (e.g.: Kassin & Fong, 1999), and others.

Deception is also studied in the context of group memberships: intergroup bias is when in a social context an individual's membership to a group takes precedence over their own identity. They can distinguish between a group they identify with, often based on certain criteria or similarities (in-group), and a group that they do not identify with based on those criteria, i.e. the ‘other’ group (out-group). Research into how group membership effects deception and attitudes around deception is essential as it has implications in day-to-day social functioning, as well as a broader impact on prejudice, stereotyping, and discrimination. Some research in this field has found that people are likely to believe their in-group members are telling the truth more often than out-group members (Fan et al., 2022), that group identification affects credibility of alibi witnesses (Rozmann & Nahari, 2021), and that people feel more guilty when deceiving their in-group members than their out-group members (Dunbar et al., 2016). Although important, intergroup deception research is scarce, and has not been a primary focus of deception researchers.

The area where research has focussed the most is interpersonal deception. An average person can be lied to as much as 200 times in a day (Meyer, 2010). People tell a lie once on average in every three social interactions they have (DePaulo, 2010) and lie thrice on average within a 10-minute conversation with a stranger (R. Feldman, 2009). A survey showed that about one third of Americans provide false information on their resume (Resume Builder, 2021), while another showed that every fourth American considered it perfectly fine to deceive an insurer (Laycock, 2018).

That said, commenting on frequency of deceptive behaviour as an average across a sample of people can be misleading as people’s propensity to lie is not normally distributed (T. R. Levine & Knapp, 2018). In fact, most lies are told by a small proportion of any sample: more than half the adults surveyed in the United States reported not lying at all, while over half the lies could be attributed to five percent of those surveyed (Serota et al., 2010). A meta-analysis found that while people lie 1.61 times on average during a 24-hour period, less than one percent of their sample told 38.5% of the lies, therefore skewing the results, and that most lies were told by people who reported being good liars (Verigin et al., 2019). This shows that there are fundamental differences between people when it comes to deceptive behaviour –

production and detection, and as such, mapping what the characteristics of deception are and what factors affects the ability to produce as well as detect lies has been the focus of a lot of research in this area.

1.1.3. Characteristics of deception

Most of the early research in deception investigated the behavioural characteristics of deception production. One of the first comprehensive summaries of research looking into cues to deceptive behaviour was by Zuckerman et al. (1981) who successfully identified 19 behaviours that were thought to be indicators that an individual was lying such as: speech errors, speech hesitancy, greater response latency, negative affect. Subsequently, more cues to deceptive behaviour and lying have been recognised. However, recent meta-analyses have found that most of these cues do not correlate with deceptive behaviour (DePaulo et al., 2003; Sporer & Schwandt, 2007) and research that has found such correlations may suffer from methodological issues (Luke, 2019). Therefore, as Hartwig and Bond (2014, p. 661) say, "there is no Pinocchio's nose—a behavioural sign that systematically accompanies deception".

This could be because there is a variability in the degree to which these cues to deception are presented by different senders. Not every sender shows the same cues to deception and it is suggested that individual differences in how consistently the senders 'leak' these cues may make them either easy or hard to catch. The concept of *cue leakage* was first introduced by Ekman and Friesen (1969) who suggested that even though senders try to control their behaviour, deception-related emotional cues leak out non-verbally as senders who are lying experience heightened emotions (such as guilt) in comparison to truth-tellers, and these leaked cues help receivers detect lies.

Zuckerman et al. (1981) furthered this concept of leakage by suggesting that any deception related behaviour, verbal or non-verbal, was affected by four factors: control (of behaviour, such as speech modulation), emotions (such as guilt or anxiety), arousal (physical, such as pupil dilation), and cognitive effort (related to creating and maintaining deception, such as coming up with a plausible lie). Again, as these factors are associated with certain non-verbal behaviours, a larger effort to control for them is counterproductive and manifests as leaked deception cues that in turn help the receiver catch the lie. Burgoon and Buller’s (1996) Interpersonal Deception Theory (IDT) also supports this theory, and extends it by proposing that both senders and receivers are leaky and deception is a dynamic interactive process. According to the IDT, senders operate strategically to control their behaviour but unintentionally leak cues, while receivers pick these cues up but further leak their own cues that indicate suspicion, which the sender in turn receives and adjusts their deception strategy. Accordingly, whether or not deception is successful is dependent on who between the sender and receiver is better at picking up cues. This suggests that the success of deception, or if investigated from the receiver’s perspective, deception detection, depends on variance in both sender and receivers. This makes sense because there are always at least two people involved in any deception: the sender and the receiver, and therefore the success or failure of deception technically cannot solely depend on the characteristics of just one.

1.1.3.1. Sender characteristics that affect deception success

Deception could be considered successful if, from a sender’s perspective, they produce a deception that the receiver cannot detect. This success is dependent on a number of sender related factors. Bond and DePaulo (2008) found that one of the biggest factors affecting veracity judgement was sender credibility, which Levine (2010) rechristened as *sender demeanour*. Demeanour can be explained as how honest or dishonest an individual sender appears (regardless of whether they are lying) and plays a very big role in whether deceptive behaviour was judged accurately. Moreover up to 98% of the variance in accuracy could be explained if the ones detecting the lies were briefed about the demeanour of the sender (T. R. Levine et al., 2011).

Another related factor affecting deception judgement is *sender transparency*. A transparent sender is essentially someone who shows a high cue leakage. Conversely, those who do not show as many cues or are not as consistent are called 'non-transparent' liars (T. R. Levine, 2010). Transparency can also be explained in terms of demeanour: someone whose demeanour matches their veracity is seen to be a transparent sender, making their veracity easy to judge, while someone whose veracity and demeanour are mismatched would be a non-transparent sender, making them harder to judge. For example, a transparent liar would be someone who looks dishonest and lies and a transparent truth-teller would be someone who appears honest and is truthful. Equally, a dishonest-appearing truth-teller and an honest-appearing liar would be considered non-transparent (T. R. Levine et al., 2011).

The fact that on average people are good at detecting lies only 54% of the times (C. F. Bond & DePaulo, 2006) has been credited to the idea that sender characteristics like demeanour and transparency can explain a large proportion of variance in deception success. However, considering that there are good lie detectors and poor lie detectors (e.g. Mann et al., 2004), there have to be individual differences in receivers that contribute to deception success as well.

1.1.3.2. Receiver characteristics that affect deception success

Since the receiver is the one who is making the veracity judgements, whether or not the deception is successful is also dependent on a number of receiver related factors. *Truth bias*, defined as a receiver's tendency to presume that the sender is being truthful regardless of their veracity (T. R. Levine et al., 1999), plays a role in deception detection accuracy and has been seen to be a factor explaining variability in veracity judgements (C. F. Bond & DePaulo, 2008). Truth bias is independent of distractions – people misremember false information as true regardless of the presence or absence of cognitively distracting tasks – and is not affected by the actual proportion of messages that are truthful or otherwise (Pantazi et al., 2018). The truth bias has been confirmed by multiple studies, e.g. Van Swol et al. (2015) found that 70.8% of receivers deemed senders to be telling the truth. Having a bias towards believing senders to be generally truthful usually serves well for people in real life because people on the whole are honest more often than they are deceptive (Serota et al., 2010, 2021), but it makes them more susceptible to being deceived. This raises the question if it can really be called a *bias* if, in real life, people encounter more truths than lies, and this propensity to believe people to be generally truthful in fact steers receivers into making correct judgements most times. If variability in veracity judgements in deception studies is affected by a truth bias and manipulating the truth-lie ratios in a deception detection paradigm can affect task performance / accuracy (T. Levine et al., 2014), it is possible that the 'bias' is an artefact of how deception detection is studied rather than a factor affecting deception detection skill at large.

Apart from construct-specific characteristics like truth bias, past research has also found other factors that affect deception detection. Receivers' familiarity with the sender has been seen to improve lie-detection (Brandt et al., 1980; Feeley et al., 1995). Receiver gender has also been found to influence deception detection: women are more accurate in detecting deception than men (G. D. Bond et al., 2005; McCornack & Parks, 1990; Tilley et al., 2005). Deception detection accuracy has also been found to vary with age, albeit with mixed results: Sweeney and Ceci (2014) found younger adults (18 to 23 years) were better deception-detectors than older adults (60 to 93 years), however Bond et al. (2005) have found that older adults are better at discriminating truths from lies.

An extensive meta-analysis Aamodt & Custer (2006) concluded that the individual differences of the receivers such as education, experience or confidence were not associated with the receivers' deception detection accuracy. They also found that accuracy rates of professionals who rationally are supposed to be better at catching lies, such as law enforcement personnel and psychologists, were not always different from that of lay people. Regardless, it is possible that instead of biological, demographic or behavioural aspects, cognitive mechanisms underlying deception are what drive its production and detection.

1.1.4. Cognitive underpinnings of deception

While the behavioural aspect of deception has been studied at length in the last few decades, the investigations into cognition of deception are a relatively recent advance. Walczyk, Roper, Seemann & Humphrey (2003) define the process of deception as twofold: the decision to lie, followed by construction of the lie, thus suggesting that a sender must suppress or inhibit the truth in order to be able to lie successfully. Inhibitory control therefore plays an important role during deception (Debey et al., 2015; Vartanian et al., 2012). Holding in mind the content of the truth they ought to be suppressing and the content of the lie is equally important; therefore, a good working memory is also essential for deception (Vartanian et al., 2012). Executive functions, overall, are thought to be essential to deception (Gombos, 2006).

1.1.4.1. Deception detection and the sender's cognitive load

The idea that deception requires multiple cognitive processes is further supported by the fact that there is usually more mental taxation involved when lying (Patterson, 2009; Vrij &

Ganis, 2014). Furthermore, increasing the cognitive load of the sender is considered to be one of the most effective ways to improve detection of lies during interrogations (Vrij et al., 2017), the extra load of undertaking other cognitively laborious tasks alongside the effort of monitoring and controlling behaviour is thought to cause leakage of even more cues (Vrij et al., 2006).

One way to increase cognitive load is to require senders to deliver their message in reverse chronological order – Vrij et al. (2008) found that having mock suspects recount a staged event in reverse order led to them leaking more cues associated with deception and this further led to increased successful detection when police observers judged the recorded recount. Similar results were found when senders were explicitly asked to maintain further eye contact, the rationale being that people tend to look away when recounting something from memory to manage cognitive load, and forcing senders to maintain eye contact increases cognitive load (Vrij et al., 2010).

Contrarily, Verschuere et al. (2018) found in their meta-analysis that increasing cognitive load decreased the difference between the reaction times of delivery of truths and lies, thus suggesting that the added load hampered working memory and suppressed the quick ability to provide true answers. However, they did find a significant difference between the reaction times of delivery of truths and lies in general (without additional cognitive load), which goes to support the idea that deception requires more cognitive effort than truth-telling does.

1.1.4.2. Deception as a cognitive developmental milestone

Another avenue of support for deception to be a cognition heavy process is that it is considered a milestone in cognitive development of children (Talwar & Crossman, 2011). Sodian et al. (1991) advocate that deception production and detection are abilities that develop throughout childhood and are of great importance in typical development; particularly, deception in children has been linked to the development of their theory of mind (ToM) or mentalizing – the ability to attribute mental states (such as beliefs) to others.

To be successful as a deceiver, a sender needs to correctly evaluate their own mental state, and the mental state of the person they intend to deceive. In other words, a sender needs to know what the other person already knows before they can formulate a successful lie. Once they have a sufficient understanding of the extent and nature of the other's knowledge and belief, they can construct a fabricated statement (or false-belief), which is different to the actual reality of the situation (viz. true belief). In the same vein, detection of deception involves the receiver interpreting the sender's behaviour and understanding that the sender intends to instil a false-belief in the receiver (Sip et al., 2008). Understandably then, mentalizing, which is the ability to represent other people's mental states to predict and understand their behaviour, is likely a major underlying cognitive process involved in producing and detecting deception.

1.2. Mentalizing and deception

1.2.1. What is mentalizing?

Imagine sitting with a friend at a café. After a little while, your friend starts fidgeting and furtively glancing at their watch. It would be natural for you to assume that either they have a prior commitment they need to get to or that they are tired / bored but do not want to offend you by cutting the meeting short. In response, you find a tactful way to end the meeting.

Sometimes we infer intentions and desires from other people's behaviour, as in the example above, and sometimes it is the other way around – we predict people's future behaviour from what we understand of their thoughts and feelings. For example, if you knew that your friend finds the café you suggest too expensive, you can predict that they would cancel the meeting to save face, therefore you can pre-empt it and say it is your treat or suggest a different café.

In both these scenarios, your assumptions help you adjust your behaviour and navigate the social scenario successfully. This intuitive ability to understand the behaviours, decisions and choices of people around us is called mentalizing and is required for smooth navigation in social contexts. Mentalizing (also known as theory of mind or mind-reading) is the process through which one attributes mental states, such as desires, feelings, intentions, beliefs, and so on and so forth, to another individual, and uses this understanding of the other’s mental state to effectively perceive, interpret and predict their behaviour (Allen et al., 2003).

Uta Frith, one of the pioneers who used mentalizing to understand the neurodevelopmental condition of autism, defines mentalizing, or Theory of Mind (ToM) as:

> The ability that we all have, we human beings, to understand other human beings – not in terms of what they behave and what they do, but in terms of what they feel and think. So, totally invisible things in your head but we have a theory of what happens in your mind. (Serious Science, 2016)

We use mentalizing in all areas of our lives; for example, it helps one realise when someone is making a joke or being sarcastic, understand that another’s feelings, desires and opinions might be different from one’s own, and recognize that the other person might have a different understanding of the situation to one’s own.

Mentalizing requires one to take on the cognitive perspective of another agent. This is known as first-order mentalizing or ToM-1. One uses second-order mentalizing or ToM-2 when they have to take the cognitive perspective of another agent who is in turn attributing mental states to someone else. One widely used way to study mentalizing is the false-belief task (FBT), which tests a person’s understanding that the information they have is not available to the agent and therefore the agent holds a belief that is different from both their own belief and actual reality (Battistelli & Farneti, 2015). A classic ToM-1 FBT would involve presenting the subject with a scenario in which the primary agent develops a false-belief about the location of target object: X puts target object in Location 1, X leaves room, Y moves target object from location 1 to Location 2, X re-enters room. The question prompting mentalizing is then asked: 'where would X look for the target object?'. The famous Sally-Anne task (Wimmer & Perner, 1983) is an example of such a mentalizing paradigm. A ToM-2 FBT would involve the subject being presented with a scenario where the primary agent develops a false belief about the secondary agent's belief: X and Y know Fact 1, X learns of Fact 2 that replaces Fact 1 in absence of Y, Y learns of Fact 2 as well but X does not know that. The question prompting second-order ToM is then asked: 'What does X think Y knows?' or 'How does X think Y will behave?'. Perner & Wimmer's (1985) ice-cream van task is an example of a second-order false-belief paradigm.

1.2.2. How are deception and mentalizing related?

Mentalizing plays a central role in deception, both in its production – where the sender needs to first understand the true beliefs held by a receiver and attempt to implant a false-belief, and in its detection – where the receiver needs to be capable of recognizing that the sender holds a different belief than he is trying to present to the receiver. In fact, mentalizing and deception are so interlinked, that early ToM studies used deception as a real-world measure of children's ToM development (e.g.: M. Chandler et al., 1989; Hala et al., 1991; Sodian et al., 1991).

Most of our present understanding of the relationship between mentalizing and deception, specifically the production of deception, comes from the child literature. One of the most common deception production paradigms in children is the temptation resistance paradigm (TRP; e.g.: Talwar & Lee, 2002) – children are told not to peek at something (usually a toy hidden under a box or cup) while the experimenter steps out and are recorded covertly. On the experimenter's return, the child is asked if they peaked inside the cup, which then prompts the deceptive behaviour. This paradigm has been used to confirm the direct relationship between deception production and mentalizing – children's deceptive behaviour was predicted by their awareness that the other person does not have the same knowledge as them of whether something is true or not (Leduc et al., 2017; Ma et al., 2015). Both first and second order mentalizing have been linked to deception production in young children: Talwar and Lee (2008) found that children's initial lies were linked to ToM-1, but their ability to continue their deception under repeated questioning was related to their ToM-2.

Mentalizing has also been linked to instances where children have developed deception skills as part of the testing paradigm. Their ToM scores predicts how quickly children pick up deceptive behaviour (Ding et al., 2018), and direct ToM training causes children to lie more often compared to children who receive alternative training about physical concepts (Ding et al., 2015), thus demonstrating a causal link between mentalizing and deception production in children.

1.2.3. Change in mentalizing and deception with age

1.2.3.1. Development of mentalizing

Until recently, it was widely believed that ToM developed at around four years of age. There is thought to be a conceptual shift at around that age, which enables children to pass ToM-1 FBTs (for a full review, see Wellman et al., 2001), and to pass ToM-2 FBTs by age five or six (see S. A. Miller, 2009 for review).

However, children as young as 15 months are thought to demonstrate mentalizing in spontaneous, non-verbal ToM tasks (Onishi & Baillargeon, 2005; Southgate et al., 2007; Surian & Geraci, 2012; see Baillargeon et al., 2010 for review). This lends credence to the idea that children younger than four years fail these tasks not because they lack ToM skills, but because they have not yet developed the worldview to comprehend the scenario or the language skills to understand the task (Bloom & German, 2000). Therefore, it is possible that children under a certain age fail *explicit* FBTs (i.e. when prompted to think about and/or explain what another thinks). When the tasks are made more *implicit* (spontaneously representing another's thought and predicting their behaviour without deliberate reflection on their mental state), ToM abilities can be observed in younger children as well. Therefore, mentalizing processes can be separated into *explicit mentalizing* ability, which develops slightly later in life, is more flexible, but is also more resource-demanding, and *implicit mentalizing* ability, which is more cognitively efficient, involuntary and inflexible (Apperly & Butterfill, 2009).

Although children successfully complete mentalizing tasks and are considered capable of attributing mental states by around 4 to 5 years of age, mentalizing capabilities continue to develop through adolescence and early adulthood (Blakemore, 2012). Devine and Hughes (2013) went beyond standard mentalizing tasks such as FBTs and tested how one applies their understanding of other's beliefs and intentions to explain behaviours encountered in everyday life by using silent films and found that participants' performance improved with age in 8 to 13 year olds. This was also substantiated by Im-Bolter et al. (2016) , who found similar results on the Strange Stories task (Happé, 1994) where the younger age group (mean age 8 years) performed worse than the adolescent group (mean age 12 years). Moriguchi et al. (2007) found that activation in the medial prefrontal cortex (MPFC) – which has been linked to ToM (Amodio & Frith, 2006) – was significantly linked with age in 9-16 year old participants as they passively watched silent animations depicting triangles (Abell et al., 2000) showing either random behaviour or mentalizing-related behaviour.

The network of brain regions associated with ToM show changes in the pattern of activation throughout development (for review, see Blakemore, 2008). However, only a few studies actually investigate the trajectory of mentalizing development beyond childhood. One of the biggest issues of testing ToM in adolescents and adults is creating a paradigm that does not produce ceiling effects. Dumontheil et al (2010) presented the Director Task (Keysar et al., 2003), a task that even adults find challenging, to participants aged from 7 to 27 years, and discovered that performance improved with age. That said, the director task has been criticized for not being a pure mentalizing measure (Rubio-Fernández, 2017). Klindt et al. (2017) circumvent issues regarding appropriateness of paradigms by instead using a battery of popularly used mentalizing tasks adapted for online administration. They tested a huge sample of 10 to 85 year old in a massive crowdsourced online study and found that mentalizing ability shows a significant improvement with age in adolescence and early adulthood across all tasks. Nevertheless, even though this study boasts of a substantial sample size across ages for the overall study – thirty thousand participants – their sample size in the youngest age groups are considerably smaller, especially when broken down by task. Concerns of smaller sample sizes for certain age groups notwithstanding, Klindt et al.(2017) provide compelling evidence demonstrating development of mentalizing beyond early childhood.

1.2.3.2. Development of deception

While development of mentalizing has been the focus of some research, development of deception beyond childhood has been investigated very little. In childhood, deception is seen to follow a pattern similar to that of mentalizing. Talwar and Lee (2002) employed a temptation resistance paradigm and found that while most 3 year olds did not lie, most 7 year olds did, albeit not skilfully enough to be successful in their deception. In a different paradigm, Sodian (1991) found that 3 to 4 years olds were capable of sabotaging a competitor by physical means if necessary, but they found it difficult to either lie or point deceptively in the wrong direction (see also: Peskin, 1992). This resonates with ToM development: 3 year olds have not developed mentalizing therefore do not attempt to deceive; while 7 year olds can pass FBTs, their ToM skills are still developing, thus although they attempt to deceive, they are not successful.

While parallels can be drawn between development of mentalizing and deception production, it is, in fact, only recently that at the relationship between development of mentalizing and developments of deception production was investigated. Zhao et al. (2021) found that children's ToM development predicts the development of how often children deceived in a longitudinal study over eight months (three test sessions, each four months apart). Children between the initial ages of 39-51 months were asked to hide a sticker under one of the two cups and mislead the experimenter so they could not find the sticker. It was found that instances of lying increased at each of the subsequent time points, and this was predicted by their understanding of diverse desires, i.e. their understanding that their own desires and someone else's desires could be different.

Like production, children's understanding of deception and their ability to detect deception evolves from age 3 to 5 years: Nancarrow et al. (2018) found a significant correlation between age and deception detection skills, as well as ToM abilities and deception detection skill, demonstrating that not only is deception-detection related to age, it is also associated with ToM development. As children get older, they are able to better judge the truthfulness of a person by examining the consistencies between the verbal and non-verbal affect (Rotenberg et al., 1989).

Beyond early childhood, literature on development of deception is scarcer than that of development of mentalizing. Of the few studies that do look at deception production as a function of age in older children and adolescents, majority focus on attitudes and perception regarding deception rather than actual ability. Jensen et al. (2004) found that young adults (college students) considered lying to parents less acceptable than adolescents did and Feldman et al. (1999) found that older adolescents were judged to be better deceivers than younger adolescents were. That said, one study did investigate development of deception production ability directly and found that in a group of 8- to 16-year-old, older participants were less likely to lie versus younger participants although this was attributed to the older adolescents understanding that the chances of successfully deception within the experimental scenario were low (Evans & Lee, 2011). Development of deception detection beyond childhood, to the best of my knowledge, has only been investigated once 40 years ago: DePaulo et al. (1982) found that older adolescents and young adults can distinguish feigned affect from genuine affect better than pre-adolescent children can.

It is therefore evident that there is a significant gap in our understanding in how deception abilities change and develop as a function of age, and if it follows a developmental pattern similar to what mentalizing has been found to have in the few studies investigating it. Considering that both mentalizing and deception are considered essential to social functioning and important developmental milestones, it further raises the question of how individuals who have difficulty mentalizing, such as those on the autistic spectrum, are affected when faced with deception.

1.3. Autism Spectrum Condition

1.3.1. What is Autism?

Autism Spectrum Condition (henceforth: autism) is a neurodevelopmental disorder that is pervasive, and is characterised by difficulties in communication and social interaction, and the presence of restricted, repetitive and stereotypic patterns of activities, behaviour and interest (Diagnostic and statistical manual of mental disorders: DSM-5, American Psychiatric Association, 2013).

About one in every 100 individuals is diagnosed with autism and males are four times as likely to be diagnosed as females (Zeidan et al., 2022). Prevalence of autism has seen a significant rise in the last few decade – 787% between 1998 and 2018 – with the rates of diagnosis showing a steeper increase in females versus males (G. Russell et al., 2022). This is attributed to better understanding of autism in the general population along with a substantial increase in reporting and diagnosis, rather than an incline in actual incidence (G. Russell et al., 2022). As of 2013, it cost £1.4 million to support an autistic individual throughout their lifespan, the amount increasing to £1.5 million if the autistic individual also had intellectual disabilities (Buescher et al., 2014). These costs are not limited to just healthcare; other factors include special education costs, costs of accommodation, of informal care and loss of caregivers' employment, and costs due to unemployment of autistic individuals (Rogge & Janssen, 2019).

The causes of autism are still under much debate and investigation. There is evidence from twin studies that autism has a genetic basis. A monozygotic twin of an autistic individual is much more likely to have high autistic traits (i.e. show some type of autistic phenotype) than a dizygotic twin of an autistic individual – over 90% concordance rates in monozygotic twins versus under 10% concordance rates in dizygotic twins (Bailey et al., 1995; Nordenbæk et al., 2014). It is further strengthened by rare genetic conditions that have high prevalence of autism, such as Fragile X Syndrome: 30 percent of individuals with Fragile X Syndrome are also diagnosed with autism (L. W. Wang et al., 2010). In fact, it is widely accepted that autism is caused by an interaction of multiple genes, although no specific genetic markers for autism have been identified yet (Abrahams & Geschwind, 2008).

Considering autism is caused by different permutations of many genes i.e. genetic heterogeneity (Rylaarsdam & Guemez-Gamboa, 2019) and that it is diagnosed based on multiple combinations of behavioural symptoms – which might differ from person to person i.e. behavioural heterogeneity (Ure et al., 2018), one prevailing theory is that autism can be explained by common atypical cognition instead (Rajendran & Mitchell, 2007). As such, different cognitive theories of autism have attempted to explain the autistic symptoms and characteristics in the context of cognition.

The Executive Dysfunction theory proposed that autistic individuals have domain-general deficits in a variety of executive functions (EF). The theory evolved from the observation that autistic individuals show certain EF dysregulation such as issues with attention switching and inhibitory control (e.g.: Ozonoff et al., 1991) that were akin to those observed in patients with brain lesions (Rajendran & Mitchell, 2007). The Weak Central Coherence (WCC) account of autism, on the other hand, suggests autistic individuals have difficulty in global processing – i.e. processing information by zooming out and absorbing the overall gist – that explains their fixation with acute details (U. Frith & Happé, 1994).

While the Executive Dysfunction theory and the WCC account are both domain general, the Theory of Mind (ToM) deficit hypothesis of autism attempts to explain the social-communicative difficulties in autism. According to the ToM deficit hypothesis, autistic individuals may struggle with representing others' mental states, i.e. mentalizing, which is the root cause of the difficulties autistic individuals have in social functioning (Baron-Cohen et al., 1985).

1.3.2. Autism and mentalizing

It has been proposed that the social-communicative deficits in autism stem from an difficulty in mentalizing, and that autistic people are not able (or less able) to attribute mental states, such as beliefs, to others in order to predict their behaviour (Baron-Cohen et al., 1985). This is supported by neuroimaging studies with autistic participants showing atypical activation in mentalizing related regions in the brain (e.g.: Happé et al., 1996) and the fact that autistic children fail or underperform in comparison to non-autistic (NT) participants in a variety of different ToM tasks in different modalities (for review, see Baron-Cohen, 2000a).

However, older autistic children and autistic adults with normal intelligence frequently pass false belief tasks (FBT) much like non-autistic people (Bowler, 1992; Ozonoff et al., 1991; Peterson et al., 2007). Scheeren et al. (2013) found that performance on second order ToM tasks surprisingly did not differ between autistic and non-autistic participants with intelligence in the normal range when they read social stories and gave reasons for the characters' behaviours. Moran et al (2011) similarly reported that the autistic participants in their study performed as well as the non-autistic participants on FBTs, but they failed to understand the intentions behind accidental harms. Consequently, the autistic participants judged accidental harms as morally wrong as attempted harms, thus demonstrating that autistic people show a limited mentalizing ability in a more ecologically valid test of intentions and behaviour. This suggests that older autistic children and autistic adults, while capable of passing traditional FBT tasks, continue to show difficulties in real-life situations that demand mentalizing.

One possible reason for this gap between good performance in FBTs and poor performance in ecologically valid ToM tasks as well as difficulty in real-world social navigation could be that autistic individuals utilize compensatory learning techniques to explicitly solve FBTs, and might still show limited spontaneous mental-state attribution or implicit mentalizing (U. Frith, 2004). This is supported by studies that use anticipatory-looking paradigms (initially developed to test ToM in pre-verbal children) to record implicit mentalizing, where autistic participants do not predict the agent's behaviour spontaneously like non-autistic participants do (Schneider et al., 2013; Senju et al., 2009). That said, some researchers have questioned the suitability of using anticipatory-looking paradigms to study implicit mentalizing in children (Kulke et al., 2018; Schuwerk et al., 2018) and highlighted the need for the development of better control conditions and replication of results, which are necessary before firm conclusions can be drawn regarding implicit mentalizing in autism. Anticipatory looking paradigms have been further criticized for not taking into consideration the existing variance in autistic gaze in social situations, and that the fast response window that is analysed may not be sensitive to the slower social processing in autistic individuals (Glenwright et al., 2021). Therefore, more ecologically valid measures that test real-world skills using mentalizing – such as deception – so as to bypass compensatory techniques, but that are not based on rapidly occurring gaze patterns like anticipatory looking paradigms, are required to investigate if an atypical ToM hinders social functioning in autism, and if it's effects are different in children and adults.

Considering mentalizing is thought to develop with age beyond childhood, there is a dearth of research looking at development of mentalizing in autism. While mentalizing in autism has been studied in children (e.g.: Schuwerk et al., 2016), adolescents (e.g.: S. J. White et al., 2014), and adults (e.g.: Schneider et al., 2013) separately, and also as part of the same sample (e.g.: Happé, 1994), very few studies have investigated mentalizing abilities in autistic individuals as a function of age. Second-order mentalizing appears to improve with age in autistic adolescents and adults (6-20 years), similar to non-autistic individuals (Scheeren et al., 2013). However, this study used social stories to measure mentalizing and the improvement in verbal ability with age could be a contributing factor to the improved mentalizing skills since autistic individual are thought to use the explicit verbal information as a means to decode other's minds in place of spontaneous ToM (Abell et al., 2000).

A more recent study by Moessnang et al. (2020) circumvented the possible confounding effects of language by using non-verbal animations (S. J. White et al., 2011) to test mentalizing in 7-30 year old autistic and non-autistic individuals and contrary to predictions, did not find any effects of diagnosis on performance on mentalizing task. Furthermore, their other findings were equally surprising and went against the established literature: they also did not find any difference in social brain activity as a function of age (although they did find effect of age on task performance) or diagnosis. This is conflicting with a plethora of previous neuroimaging literature showing effect of age in the social brain activation in non-autistic samples (see Blakemore, 2008 for review), of atypical activation in the social brain of autistic samples (see Sato & Uono, 2019 for review), and of cognitive-behavioural studies with autistic samples (see Baron-Cohen, 2000b for review). Therefore, it is advisable that their findings be interpreted with some caution and that the relationship between development and mentalizing in autistic and non-autistic individuals be investigated further.

1.3.3. Autism and deception

Deception production in children with autism has been a popular area of investigation, mostly because deception production is considered a marker of ToM development and a cognitive milestone. When compared to non-autistic children, autistic children find it more difficult to occlude information (vs. occlude an object) in a penny-hiding game (Baron-Cohen, 1992), to strategically point towards the wrong box to mislead the examiner (J. Russell et al., 1991), to maintain consistency in follow-up questions following an initial lie (A. S. Li et al., 2011), and to deceive in retaliation to being deceived (Yi et al., 2014).

Deception in autism beyond childhood has hardly been explored. In fact, according to a scoping review by Bagnall et al. (2022), only four studies have looked at deception production in adults, and with mixed results in regards to their success and performance in comparison to non-autistic and other clinical populations. However, none of these studies have investigated autistic individuals' deception production abilities in context of demeanour, which has been seen to be a major driver of deception success (T. R. Levine, 2010). Since autistic individuals generally make unfavourable first impressions (Alkhaldi et al., 2021) and are seen as less trustworthy (Lim et al., 2022), it is important to examine autistic people's ability to successfully deceive in the context of demeanour to gain a better understanding of deception in autism.

Deception detection in autistic adults is a similarly under-researched field and the few studies that have investigated it have not focussed on deception detection per se, but as a measure of their understanding of complex social construct alongside sarcasm and/or jokes (Mathersul et al., 2013; Warland, 2014). Only one study, to the best of my knowledge, directly investigates deception-detection ability in autistic adults (D. M. Williams et al., 2018): autistic adults were found to be less accurate in overall lie detection and were worse at detecting even transparent lies (versus non-autistic individuals). However, even though deception detection ability was investigated in relation to mentalizing ability in non-autistic participants, the study did not look at the relationship between mentalizing and deception in autistic participants.

Therefore, there is a further need to investigate deception detection in autism in relation to the development of mentalizing. This is especially true as difficulties in mentalizing and detecting deception are likely to cause difficulties in everyday life, which go beyond simple inconvenience and have far-reaching implications.

1.4. Implications of poor deception detection

Researchers of deception have often implied that everyday one-to-one lies are inconsequential as there are no large repercussions (e.g.: Hartwig & Bond Jr., 2014). However, everyday one-to-one deception too can have sweeping consequences on an individual level. For autistic people, who already face difficulties in social interactions, the further effects of not being able to judge when they are being lied to can be debilitating.

Bullying and social rejection is at its peak during adolescence (Nansel et al., 2001), and weaker mentalizing has been found to predict bullying victimization in adolescents (Shakoor et al., 2012). Autistic adolescents are more prone to being bullied than their non-autistic peers: autistic adolescents report a prevalence of 45% for peer victimization verses 10% in non-autistic adolescents (Sterzing et al., 2012). Considering that autistic individuals are thought to have difficulties in mentalizing, this is likely to be a contributing factor to their vulnerability to victimization. Deception detection may be a factor that modulates the relationship between mentalizing and vulnerability to victimization, since most bullies are also known to be good deceivers, e.g. engaging in instilling false beliefs to perpetuate rumours (Sutton et al., 1999a). The fact that autistic individuals who experience such victimisation are also at greater risk of mental health difficulties (Cappadocia et al., 2012), makes this issue even graver.

Therefore, it is imperative to investigate the relationship between mentalizing and deception-detection, the development of these abilities with age, and their association to bullying, victimization and mental health issues in both autistic and non-autistic populations.

1.5. Conclusions

In this chapter, I underscored how prevalent deception is in our everyday lives and summarised what we know about deception's cognitive drivers – especially mentalizing, and how mentalizing and deception develop hand in hand. More so, I shed light on the gaps in our understanding of both these mechanisms – the paucity of research looking at development of mentalizing and deception beyond childhood, especially so into adulthood. I introduced how an apparent difficulty in mentalizing is thought to be a contributing factor to the social-communicative difficulties that autistic individuals face, especially their ability to produce and detect deception, while highlighting the scarcity of deception research in autism, especially in adults. Finally, I brought home why this is important – our understanding of how deception is detected, how this process changes over time, and how this ability is linked to mentalizing will likely help us understand (and therefore take measures to counter) the peer victimization that autistic individuals face, which may partly be due to their difficulties in deception detection.

In the following chapters, I will try to answer the questions that have been raised throughout this chapter. In chapter 2, I detail how I created a novel deception paradigm that I use throughout my PhD. I discuss the limitations of past deception studies and explain how I sought to counter these in my novel paradigm, along with detailing a step by step process of building a deception production paradigm, producing stimuli, and creating a viable deception detection paradigm.

In chapter 3, I look at factors that affect deception production. I test if senders' deception production ability remains constant when lying about different things, and if senders choose the less cognitively taxing option when trying to deceive. Further, I investigate if successful deception is driven by sender demeanour. I also examine if autistic traits in non-autistic adults affect overall deception success and if mentalizing too plays a role in how successful they are at deception production.

In chapter 4, I tackle the central question of this thesis: how mentalizing and deception detection develop from late childhood, through adolescence, to early adulthood in both autistic and non-autistic samples. I investigate if there is an effect of age on mentalizing, and if it interacts with diagnosis. I attempt to extend Williams et al.'s (2018) findings, demonstrating weaker deception detection in autistic adults (versus non-autistic adults), to a younger sample and furthermore, directly examine the relationship between mentalizing and deception-detection in autism. Finally, in this chapter, I also look at whether mentalizing and deception detection affect peer victimization and mental health in autistic and non-autistic individuals.

In Chapter 5, I venture into a different line of inquiry regarding other factors that may cause autistic individuals to struggle with deception detection – specifically group membership. It is here in chapter 5, that I introduce the double empathy problem (D. E. M. Milton, 2012) which is quickly gaining traction in the autism research community as an alternative explanation for why autistic individuals face the social difficulties that they do.

Finally, in chapter 6, I bring it all together to summarize my findings and their implications. I showcase how my research builds upon our existing understanding of deception, mentalizing, and autism. Finally, I discuss the limitations of the current research and suggest future avenues for autism and deception research.

Chapter 2. Novel Deception Paradigm

2.1. Introduction

Deception is such a common occurrence in our daily lives that people have been developing ways to detect lies for millennia. According to Vicianova (2015), people in Ancient China (circa 1000BC) would get the accused liar to hold some dry rice in their mouth and then spit it out. If the rice was found to be dry after expelling, it was taken as a confirmation of guilt. It was based on the idea that anxiety induced by the fear of being caught would cause the accused's mouth to go dry. Although logically sound, it was not a very effective means of detecting deception – unbeknownst to people of that era, dry mouth could be caused by a multitude of reasons including depression, panic disorders etc. making this technique of lie detection unreliable and invalid.

We have come a long way since then and our understanding of deception has grown exponentially over the last few decades, as have our ways to detect deception. Deception has been studied extensively and consequently, paradigms exploring deception have ranged from simple laboratory experiments to elaborate staged experiments mimicking real life scenarios. While they have helped us understand deception – production and detection – better in their own way, they have also brought to light some shortfalls of how deception is studied.

In this chapter, I am going to shed light on some of the shortcoming of previous deception paradigms / studies, introduce the novel tasks used throughout this project and discuss how they attempt to mitigate some of these shortcomings. I will expand on how these tasks were developed, and finally, I will discuss the strengths and possible limitations of the novel tasks.

2.1.1. Issues with past deception detection studies

A large proportion of deception studies take place in the laboratory, and as such, they have been criticised for lacking ecological validity and generalisability outside of the laboratory. Two primary areas of limitation are how the stimuli are produced i.e. how the lies that are used in the deception detection paradigm are generated, and how the actual deception detection tasks are structured.

As explained in Chapter 1, deception is a social activity and there are always two people involved – a sender and a receiver. Therefore, deception detection and production are two sides of the same coin and design issues related to one will have an effect on the other. For example, how lies are produced will have an effect on the validity of any deception detection paradigms using those lies and how lies are detected will feedback into how the senders' deception production is assessed. The limitations mentioned below therefore may be relevant to one or both processes, regardless of whether the shortcomings are with production or detection stages.

2.1.1.1. Producing deceptive stimuli

2.1.1.1.1. Choice in lying – self-selection

One of the most valid criticisms of laboratory-based lie production experiments is that instructing senders on when to lie and to tell the truth is not ecologically valid: in the real world outside laboratory conditions, people chose when and if they want to lie. Some studies randomly assign senders into lying or truthful conditions. For example, a study looking at the skill of detectives at detecting deception used videos of confederates who were either instructed to tell the truth or to lie (Vrij, 1993) or a mock job interview study investigating the vocal stress during deception randomly assigned participants into truth telling or deception conditions (O'Hair et al., 1990). Other studies ask participants to prepare both truths and lies, and instruct them when to say which, for example the false opinion paradigm used by Wright et al. (2012) and Noguchi and Oizumi (2018).

Varying brain activation patterns have been observed when senders are instructed to lie or tell the truth, in comparison to when they are allowed to spontaneously choose and decide for themselves whether to deceive or not (Yin et al., 2016). The difference between reaction times when telling the truth and when lying decreases when participants choose to lie instead of being directed to lie, thus showing that having a choice lessens the cognitive load involved in lying (E. J. Williams et al., 2013). As additional cognitive load causes senders to leak more cues (Vrij et al., 2006), it is important to control for any additional cognitive load that and experimental paradigm might introduce by taking away senders' choice. Furthermore, choice is also important because people might be aware of their own capabilities as liars and may choose to lie or avoid lying accordingly (Frank, 2008; T. Levine, 2010; T. Levine et al., 2010), thus forcing senders to lie when they would likely not lie in real life is not a true reflection of their deception ability. Additionally, it has been seen that choice increases intrinsic motivation because it increases the feeling of autonomy and being in control of one's actions (Patall et al., 2008). Therefore, a sender having a choice between lying or not, on how or when they want to lie, could greatly influence the qualities of lies. This is therefore also an important factor when trying to control for the quality of the stimuli in a deception detection task.

2.1.1.1.2. Low motivation

People outside the lab choose to lie because lying solves a problem that telling the truth would not, i.e. they are motivated to lie as lying aligns with their communication goals (T. R. Levine, 2018). However, in a laboratory setting, senders do not have the incentive to lie as there is often no positive or negative consequence (Ganis & Keenan, 2009). In fact, Bond and DePaulo (2006) found that in over half the studies they investigated in a meta-analysis senders in the laboratory, unlike real life, did not have any motivation to deceive, or to deceive *successfully*. This might explain why neither were the lies in these studies ecologically valid nor their success rates generalizable (DePaulo et al., 2003). Certainly, when people deceive in real life, whether it is white lies or malicious lies, they want the recipient to believe them; therefore, they are more invested in the deceptive act.

This lack of incentive or motivation can have consequences on the believability of the lies. According to the Motivational Impairment Hypothesis (DePaulo et al., 1988; DePaulo & Kirkendol, 1989), highly motivated senders are more likely to leak cues and their truths are more easily discernible from their lies. This leads us to believe that the low motivation in lab settings (vs real world) could make the lies especially difficult to judge, and while senders might be better at controlling cues in low motivation scenarios (vs high motivation / high stake scenarios), this inflates their deception production success in the lab. Therefore, it is important that the sender in deception paradigms be motivated to lie, and to lie well, so that an accurate understanding of the senders’ production skills and receivers’ detection skills can be obtained.

2.1.1.1.3. Low interaction

“Lying is a cooperative act” (Meyer, 2011); outside of a lab environment lying requires at least two parties – one who is lying and one who is being lied to. However, traditional deception experiments – both production and detection – often overlook the importance of interaction between a sender and receiver. In a meta-analysis of 206 deception studies between 1941 and 2005, Bond & DePaulo (2006) recorded how interactive each of the studies were. They found that one third of the studies included were designed such that senders were allowed no interaction during the session – i.e. they either lied to a passive observer or lied in the absence of a second person (e.g. to a camera). Half of the studies did involve some interaction – but these interaction partners did not judge their lies. Only about 9% of included studies involved any interaction between a sender and a receiver, i.e. person who actively judged the deception.

As reviewed in Chapter 1, there are multiple factors related to both the sender and the receiver that affect the success of a lie. Whether or not a sender is successful in their deception is dependent on both a sender's lying ability and a receiver's gullibility. However, the *quality* of the deception can also be affected by the interactive nature of the conversation. Burgoon et al. (2001) demonstrate that a more interactive session gives senders more control over non-strategic behaviour – such as leakage of deception cues (Ekman & Friesen, 1969) – as well as adjusting their behaviour and actively regulating demeanour in response to receivers' verbal and non-verbal feedback. This makes it harder for receivers to detect deception in interactive settings versus deception in a setting with less interaction; thus, testing receivers' deception detection skills on stimuli recorded in low interaction settings, instead of an interactive setting similar to how deception takes place in the real world, gives us an inflated account of their skills. Further, it does not give us a correct measure of the senders' productions skills either as the paradigm does not allow them to adapt and better their deception skills as a response to non-verbal feedback from a receiver.

Learning from feedback from one's conversational partner is a common feature of any communication, and Miller and Stiff (1993) point out that senders in traditional deception production studies often receive very little feedback while they lie or tell the truth, and this can affect the quality of any subsequent deception. This means that for the results of a lab-based deception experiment to be remotely generalizable to real life, stimuli need to be generated within an environment that allows interaction between a sender and a potential receiver.

2.1.1.1.4. Simulating real life

Some deception experiments create scenarios where participants believe that there are real-life consequences of their deception failing and therefore they self-select lying as a problem solving mechanism, are motivated while lying, and are interacting with someone who is judging the information they are providing. These studies help mitigate the issues that have been outlined above, however these paradigms often come with their own caveats.

Usually, in order to create such high-stake, ecologically-valid (vs traditional deception experiments) studies, participants are put in a mock crime scenario (Frank & Ekman, 1997) or situations where they are tempted to cheat on a test, and then are interviewed about it (e.g. Levine, 2007). These types of paradigms usually have a premise of a transgression having taken place and have an interrogatory style – i.e. participants are asked multiple follow-up and leading questions that help the receiver, who poses as an interviewer/interrogator, detect the deceit. As cross-examining has been seen as one of the best deception detection techniques (T. Levine et al., 2010), they make for excellent deception detection paradigms but can only be generalised to situations where deception is expected or where the receiver is confrontational and/or oppositional, such as in the criminal justice system.

A large proportion of deception studies include lies about transgressions (see DePaulo et al., 2003) even though day-to-day deception does not usually take a similar form. In real life, people do not expect to be lied to, and therefore do not spontaneously detect lies unless they are prompted (Clare & Levine, 2019; T. R. Levine et al., 1999) and have a tendency to believe that the people they interact with are generally honest (Truth default state; Levine, 2014). Generally, people do not cross-examine and question every conversation they have with another person. A victim of a bullying incident where they are being deceived, for example, does not always think to question the motives of their bully and the genuineness of the information they received from the perpetrator.

The other issue with such paradigms is that senders are often prompted to either accept or deny a leading contextual question (such as an accusation). This produces lies that are very succinct and are not cognitively effortful to produce, unlike deception in the case of white lies, fraud pitches, or bullying incidents, which are usually freestyle speech that require more structure and cognitive effort (Frank, 2008).

Therefore, when trying to generate stimuli that are more ecologically valid in terms of sender choice, interaction, and motivation, it is essential to keep in mind the context of the deception detection paradigm. Day to day deception is very different from deception in high-stake scenarios such as a criminal investigation, and a deception detection study must be aligned with the ecology it seeks to generalise its findings to.

Other studies have devised creative ways to test deception that is unsanctioned and unprompted, thus preserving sender choice and motivation. Van Swol et al. (2012) for example, gave senders money that they were required to split with either a stranger or a friend, deciding how much to allocate to themselves and the receiver. As receivers were not provided information regarding the amount of money that was to be allocated, this provided the senders an opportunity to deceive the receivers about the sum by either omission (they did not mention the amount they were allocating) or by lying (they mentioned a false sum). However, more than two-thirds of the senders did not deceive at all. While this was taken as a confirmation of the veracity effect, it did not serve as an effective deception-*detection* study since most senders chose to be truthful. Therefore this shows that a study where deception is completely unprompted, while more ecologically valid, is not viable when it comes to designing deception detection studies or collecting stimuli for the same. It is necessary to find a way to preserve sender choice, motivation, and interaction, while making sure that the stimuli collected are usable in a detection paradigm.

2.1.1.2. Deception detection task design

Alongside mitigating the above-mentioned issues while generating the stimuli, i.e. the deceptive and truthful statements from senders, it is essential to understand how traditional lab-based deception detection studies are designed and how these can be improved on.

Traditional deception detection studies involve receivers watching videos of senders and making judgements about their veracity. The issue with a lie-detection paradigm like this is two-fold. One, a task where participants watch multiple people either lie or tell the truth, and have to categorize them as such, implies a truth-lie base rate of 50-50 (T. R. Levine, 2018). Unlike in the lab environment, people in day-to-day life do not lie (and therefore experience being lied to) in half of all their interactions. In fact, most people are usually honest and the majority of lies are told by a small fraction of people (Serota et al., 2010, 2021). This, along with the fact that people have a tendency to believe that the people they interact with are generally honest, i.e. people have a truth bias (T. R. Levine, 2014), means that people have a higher recognition score for true statements than for lie statements. Therefore, experimental paradigms with fixed base-rates must draw conclusions within these experimental constraints: accuracy has been seen to vary as a function of the base-rate (T. Levine et al., 2014). This highlights the fact that studies that employ dichotomous truth vs lie judgements to gauge deception detection ability are confounded by biases that are independent of veracity of sender message or skill of receiver.

The second limitation of typical laboratory deception experiments where every sender only either lies or tells the truth is the variance in sender demeanour. Sender demeanour, as per T.

R. Levine et al. (2011), is how honest or dishonest a person appears regardless of the veracity of their statements and has been observed to be one of the most important factors that drives deception detection above the receiver's deception detection ability. Controlling for demeanour across truth-tellers and liars is therefore important and one possible way to do so is by including both truthful and deceptive stimuli from each sender, thereby keeping demeanour constant across truths and lies.

2.1.2. Deception paradigm in present project

In the present project I developed two deception production tasks that sought to address some of the mentioned issues of past deception paradigms. To develop the stimuli, participants were tested in pairs, and played 'deception games'. They took turns to be the sender and the receiver, therefore making sure that senders interacted with another individual: they received non-verbal feedback during each trial and verbal feedback between the trials from the receiver who was actively judging the information they were providing rather than passively observing them.

To counter both the issue of forced instruction and lack of motivation in previous studies, these tasks aimed to giving senders a *choice* in whether to engage in deception in some way, thus providing intrinsic motivation for the sender to deceive successfully.

The goal of both the games was the same – to deceive the receiver and to successfully detect whether the sender was lying or telling the truth. In the first game, the senders described illustrated pictures to the receiver, while trying to deceive them about the contents of the picture. On each trial, while describing the picture, senders could choose to either lie about the content of the picture or they could double-bluff, i.e. tell the truth like it was a lie. The receiver in this game had to judge whether the sender was describing the picture contents untruthfully or truthfully. In the second game, participants played the popular icebreaker game “two truths and a lie” and for each trial the senders supplied two truths and one lie about themselves. The receiver had to pick out the lie from the three statements supplied. While the goal of this game for the senders was, again, to deceive the receiver on every trial, they were free to choose in which order they delivered the truths and the lie.

To further incentivize participants, both senders and receivers were offered bonus monetary rewards for every successful deception (sender) or every successful detection (receiver). Although it has been previously thought that extrinsic rewards negate the effects of intrinsic motivation (Deci et al., 1999), a more recent meta-analysis has shown that intrinsic motivation coupled with extrinsic incentive leads to better performance (Cerasoli et al., 2014). Therefore, we reasoned that the combination of intrinsic motivation of self-selecting their deception and extrinsic motivation of bonus monetary reward would maximize senders’ drive to deceive well.

I then took video recordings of the participants in the two deception production games and developed them into stimuli for two deception detection tasks. These deception detection paradigms departed from traditional deception-detection tasks in two significant ways. Rather than making a judgement on a video-by-video basis – i.e. watching one video by one sender and judging if it was the truth or a lie, receivers were shown three videos from each sender and were asked to identify which of the three was the lie. This helps counter the truth-bias, as receivers are guaranteed that one of the three statements is a lie.

Second, demeanour was controlled for between truthful and deceptive stimuli as both truthful and deceptive stimuli in each trial were provided by the same sender. This, therefore, helps us measure receivers’ deception detection capabilities independent of sender demeanour. Even then, an additional task was created to measure the perceived honesty of the senders, to evaluate the degree to which senders' deception success is dependent on their demeanour, even when it is controlled across stimuli of different veracity.

2.1.3. Aims of this chapter

The main aim of this Chapter was to create two novel deception detection tasks that attempt to overcome the limitations of past studies. As such, there were three stages in the creation of these tasks, and each stage had its own aims.

Stage 1 – lie production – involved inviting participants to produce truthful and deceptive statements that would be used as stimuli in later deception tasks. It also helped us investigate how senders' deception production abilities are affected by cognitive and behavioural characteristics such as autistic traits, mentalizing, and demeanour (Chapter 3)

In stage 2 – the baseline study – groups of receivers were invited to judge the stimuli collected in stage 1. This was done for two reasons: first, to ensure that the deception production abilities of the senders in stage 1 were more objectively judged, independent of the subjective deception detection abilities of the individual receivers in stage 1. In other words, multiple receivers were asked to judge the sender to make sure that when, for example, a sender was successful at deceiving the receiver during stage 1, it was because they were good at deceiving and not because the particular receiver in the original pair was poor at detecting deception. The second reason for the baseline study was to shortlist stimuli for any future deception detection paradigms.

Stage 3 included two rounds of pilot studies to ensure that the tasks created by shortlisting videos from Stage 1 and Stage 2 worked as expected i.e. non-autistic adult participants performed over chance on the tasks.

As this chapter primarily deals with creation of the tasks, it does not have a separate results section. Instead, findings (where relevant) from each stage are presented alongside the methods. This chapter is structured in such an unconventional way because findings from each stage feeds into the following stage.

2.2. Methods and findings

2.2.1. Stage 1 – Lie production study

2.2.1.1. Participants

Forty-two males between 18 and 26 years of age (*M* = 22.36 years, *SD* = 1.54 years) were recruited through UCL Psychology and Language Sciences online participant database. All participants were either native English speakers (*n* = 30), or were fluent in English (*n* = 12).

Participants were tested in two sessions: one paired session that included all the deception tasks and an individual session where other data such as, autistic traits and mentalizing ability, was recorded. A proportion of participants (*n* = 13) did not complete the individual session due to unavailability or scheduling conflicts, however their data from the paired session was used to develop the deception detection tasks.

Initially (pre-CoViD19) the focus of my PhD was solely on the development of deception detection in males, as males and females tend to show cognitive variation in puberty and adolescent development and are therefore usually studied separately (Blakemore et al., 2010). Furthermore, autism is diagnosed at a higher rate in males (Fombonne, 2009), making them easier to recruit. Therefore, as participants are more likely to engage with stimuli that they perceive to more relevant or similar to them, and keeping in mind the target population for the task would be used in, the participants recruited at this very first stage were all males. Similarly, the age range was restricted to young adults to make sure that individuals in the video stimuli would be young enough to be relatable to by 11-30 year olds.

2.2.1.2. Ethics

Informed consent was obtained from each participant before the commencement of the study. Participants were aware that the testing sessions were being recorded and that the recordings would only be viewed by researchers specifically linked to this project – unless additional consent was provided. Participants were given the opportunity to give additional optional consent regarding use of their video recording in various capacities such as in conferences, future research etc. Participants were made aware that this was strictly *optional* and declining would not disadvantage them in any way, or affect their participation or compensation for their time and effort. All procedures were approved by UCL Research Ethics Committee (Project ID Number: 14807/001).

2.2.1.3. Measures and Procedure

Testing was split into two parts: the deception paradigm, owing to the nature of the tasks, required the two participants to be tested together (joint session), whereas all other measures were administered individually (individual session). The individual tasks (viz. a mentalizing measure, an autistic traits measure, an IQ estimate tasks, an imagination task, and an inhibitory control measure) did not contribute to the process of developing the deception paradigms, and therefore they will not be described in this chapter. Instead, any relevant tasks will be described in further chapters.

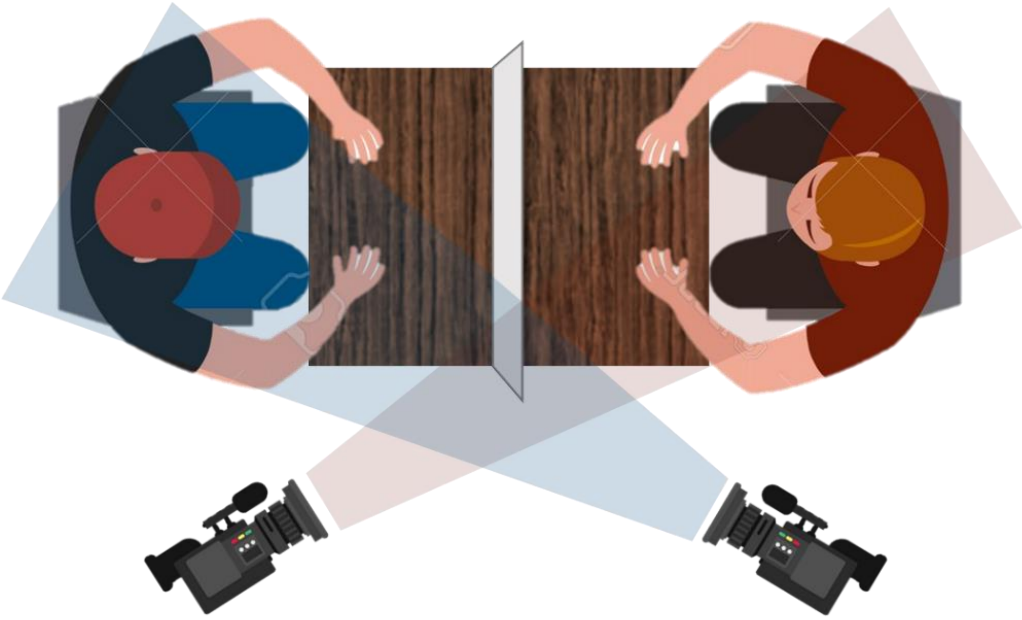
Due to scheduling constraints, the individual session took place either before or after the joint session; while most participants completed the two sessions consecutively, some participants were seen on different days. Participants were paid an hourly sum to thank them for their time and effort, plus their reward from the deception-paradigm once they had completed all the tasks. If the two sessions took place on different days, participants were paid for each session separately after its conclusion.

At the start of the joint session, the participants were provided with some basic information regarding the experiment and were asked to sign a consent form. Two Panasonic video cameras, which were used to video record the tasks, were turned on once participants had signed the consent form.

Two novel deception tasks were devised, both presented as 2-person games, where the goal of one participant was to deceive the other person. The first task was the *Weird Pictures task – Production (WP-P)* which was followed by its control condition, called *Picture Description task (WP-Control)*. The second deception task was *Two Truths and a Lie - Production (TTL-P).* For each trial (in both tasks), one person was the *sender* while the other was the *receiver*. The two participants switched roles after every trial. Participants were seated opposite each other at a table, with a 2-fold board divider placed between them, so that they could see each other but not the stimuli pictures in front of the other person. Figure 2.1 demonstrates the experimental set up.

Figure 2.1

Schematic of experimental set up during Stage 1 – Lie Production



Note. Bird's eye view: Participants were seated at a table in front of each other with a screen between them so that just shoulders and face were visible. Cameras were placed at each participants shoulder; each camera recorded the opposite participant (depicted by the red and blue beams).

2.2.1.3.1. Weird Pictures – Production (WP-P)

Illustrations from Chuck Dillon's 'That's silly' and 'What's wrong' collections from Highlights kids' magazine (Dillon, 2009, 2010) were cropped to create the picture stimuli for this task. Each picture had some 'weird' elements to them, i.e. items that were out of context. For example, one picture had a man in scuba diving gear sitting in a park fountain while a girl used his oxygen pipe as a skipping rope (Figure 2.2). Another picture depicted a boy walking a tortoise on a lead outside a building, while a cat in a tutu can be seen through a window, stretching at a barre (Figure 2.3). These illustartions were selected as they were richly detailed, making them good stimuli to describe, and included unusual items, thus making it less obvious whether the statements provided by the participant was a lie solely from the contents of their description.

Twenty such pictures were created, each 13cm x 13cm, printed and laminated. They were divided into four sets of five pictures each, one set for each participant to describe in each task. Rather than randomly allocating the pictures in four sets, I rated each picture on 'weirdness'– i.e. how out of context the various elements in the picture felt – and 'busyness' – i.e. how extensive the level of detail / number of elements each picture was – and distributed the pictures in the four sets such that the all sets had similar ratings. Each of the two participant saw one set for the WP-P task and the other set for the WP-Control task, and the sets were switched between tasks for the next pair of participants.

Participants were provided with both verbal and written instructions. They were told that they would be describing some pictures to the other person and shown an example of a picture they would be describing, which had a cowboy riding a horse in the middle of a stream alongside what appears to be a canoe race. It was emphasised that the pictures, like the example, would have some weird elements to them.

The goal of the game, as a sender, was to deceive or mislead the other person. They could do this by *lying* about what they saw in the picture and make the receiver believe that they were telling the truth or they could *double bluff* i.e. describe the picture's contents truthfully but in such a way that it appeared to be a lie. As a receiver, the participant was to pay close attention and determine if the sender was lying or describing the picture truthfully.

To help the participants keep time, bell sounds were played to signal the beginning/end of each section of a trial:

> Bell 1: Sender turns over a picture and has 15 seconds to memorise and familiarise themselves with the picture.

Bell 2: Sender turns the picture so it is face down on the table and has five seconds to strategize.

Bell 3: Sender describes the picture, either truthfully (i.e. double-bluffing) or by lying about it, for 30 seconds.

Bell 4: Sender stops speaking. Receiver is prompted by the experimenter to give their verdict about whether the sender lied or told the truth about the picture.

The sender’s deception type (lie/double-bluff) and receiver’s response was recorded by the experimenter on a response sheet. If the sender successfully misled the receiver, they won the point. If the receiver successfully guessed whether the sender lied or told the truth, they won the point. Both participants were aware of who won each round.

Participants were encouraged to speak for the full 30 seconds; however, they did not always speak for the whole duration. Participants were also made aware that they were playing for an extra incentive: they would be rewarded 50p for every point they won. They were therefore encouraged to make their lies and double bluffs sound as convincing as possible in order to win the bonus money.

There were 10 trials: each participant was a sender for five pictures, and a receiver for the remaining five pictures.

After the task, both participants were asked some debriefing questions enquiring about how they decided whether to lie or double-bluff, how they created the content of the lies, what they looked for when they were the receiver, and if they identified any ‘tells’ or deceptive cues of the other person.

Figure 2.2

Example of Chuck Dillon's 'What's wrong' illustration

[Copyright image redacted]

Note. Example of Chuck Dillon's 'What's wrong' illustration (left); cropped section presented as stimuli in Weird Picture task (right). From "Highlights for Children", by Chuck Dillon. Copyright ©2010 Highlights for Children, Inc., Columbus, Ohio. All rights reserved. Adapted with permission.

Figure 2.3

Example of Chuck Dillon's 'That's silly' illustration

[Copyright image redacted]

Note. Example of Chuck Dillon's 'That's silly' illustration (left); cropped section presented as stimuli in Weird Picture task (right). From "Highlights for Children", by Chuck Dillon. Copyright ©2009 Highlights for Children, Inc., Columbus, Ohio. All rights reserved. Adapted with permission.

2.2.1.3.2. Picture Description Task: Weird Pictures task control condition (WP-Control)

The *Picture Description Task (WP-Control)* was used as a control for the WP-P task. Participants described the set of pictures they did not previously see in the WP-P task.

Participants were given verbal and written instructions, and were told that they would see pictures similar to those in the WP-P task. Instead of misleading the other and competing against each other, participants were required to work *with* each other: the sender was asked to describe the picture in enough detail to help the receiver guess which picture the sender was looking at. The bells and durations of sections were identical to the WP-P task. As with WP-P, participants took turns being senders and receivers. However, instead of lying or double-bluffing, participants were instructed to describe the picture truthfully. The receiver was required to pay attention to the description and was presented with nine pictures (five actual stimuli pictures and four decoys) from which to pick the one that best matched the sender’s description in that trial. The receiver’s responses were recorded by the experimenter on a response sheet. As there was no winner in this game, there was no monetary incentive.

The WP-Control task was developed to be the matched ‘truth’ condition for the WP-P. Therefore, the statements recorded in this task served as the ‘truth’ stimuli in subsequent deception-production analyses and in the deception-detection paradigm. The WP-Control task was also used to establish that the senders were capable of describing pictures and did not face any difficulty comprehending and memorising the pictures, and that the receivers were capable of understanding the other person’s description. The nature of the task was such that a ceiling effect was expected. Accuracy ranged from 80-100% and no participants were deemed to show difficulties in describing the pictures as a sender or understanding the descriptions as a receiver.

2.2.1.3.3. Two truths and a Lie – Production (TTL-P)

The traditional icebreaker game 'Two Truths and a Lie' (TTL-P) was adapted for this study to elicit untrue statements and matched true statements. Participants were given verbal and written instructions. They were told that, as in the previous games, there would be a sender and a receiver and that the roles would switch after every trial. In each trial, the sender would be provided with one of four predefined topics: Weekend ("What do you usually do on weekends?"), Future ("What are some of your future plans or long term goals?"), Hobbies ("What are some of your hobbies or what do you do in your free time?"), and Summer ("What are your plans for this summer?"). Topics for both participants were the same, but were provided in different orders; these two orders were the same for all pairs. They needed to produce three statements about that topic: two of them had to be true, one had to be a lie, and each at least 30 seconds long. The order in which the truths and the lie were presented was the sender's choice. The goal of the receiver was to discern which statement was the lie. As with the previous tasks, bells were used to demarcate time:

Bell 1: Sender has 30 seconds to prepare their statements

Bell 2: Sender provides first statement for full 30 seconds

Bell 3: Sender provides second statement for 30 seconds.

Bell 4: Sender provides third statement for 30 seconds.

Bell 5: Sender stops speaking. Receiver is asked for their verdict about which statement they thought was the lie.

At the end of each trial, the sender indicated which statement was the lie. Both the receiver's verdict and the sender's response were recorded by the experimenter on a response sheet. The receiver won the point if they correctly identified the lie; otherwise, the sender won the point.

As in the WP-P task, participants were given the chance to win a monetary bonus of 50p for every point they won to encourage them to make their lies sound as believable as the two truths.

After the task, both participants were asked debrief questions about how they decided on an order for the truths and lie, and how they created the lie. They were also asked if all the truth statements were indeed true (to double-check the authenticity of their statements and avoid mismarking a situation where a sender provided more than one lie in a trial). Finally, they were asked how they distinguished the other participant's lie from their truth statements.

2.2.2. Stage 2 – Baseline study

2.2.2.1. Participants

Seventy-six males between 18 to 30 years of age (*M* = 22.64 years, *SD* = 3.17 years) were recruited through the UCL Psychology and Language Sciences online participant database in exchange for either a monetary reward at standard departmental hourly rates, or course credits. Thirty-seven of these participants were native English speakers, while 39 were fluent second language speakers of English with an average age of English language acquisition of 6.29 years. Participants took part in the study in four different groups / batches and watched only one of the four sets of stimuli. The batch-wise breakdown of participant demographics can be found in Table 2.1.

Table 2.1

Batch-wise demographic breakdown for Stage 2

Batch no.	*n*	Age *M* (*SD*)	No. of L2 English speakers
1	20	23.2 years (3.11 years)	11
2	18	23 years (2.57 years)	10
3	17	21.53 years (4.05 years)	10
4	21	22.67 years (2.73 years)	7

Note. *M* = mean, *SD* = standard deviation, *n* = sample size, L2 = second language

2.2.2.2. Measures and procedure

Each batch of participants was invited into a seminar room where they were provided with an information sheet giving them a brief overview of the study and a consent form to fill. The study took approximately 2 hours including breaks. At the very start, participants were given two questionnaires to fill out: (a) a basic demographic questionnaire; and (b) an autistic traits questionnaire (not analysed as part of this project and not relevant to the development of the deception tasks). They could fill these in either at the allocated time right at the start of the session or during the breaks. They were also handed response sheets for all the video tasks.

All video stimuli were projected on a screen and audio was played via the inbuilt speakers in the seminar room. Videos for every video task were merged in one file and were played with the aid of VLC media player. Instructions for each task were delivered verbally by the experimenter and presented on the screen.

Videos from participants who took part in Stage 1 (henceforth: senders) were shortlisted (details in sections 2.2.2.2.1 and 2.2.2.2.2 below), edited, and batched before being collated as video tasks and shown to participants in Stage 2 (henceforth: receivers).

Receivers were given some basic information about how the videos were filmed – senders took part in a study in pairs; they did not know each other; senders played deception games, attempting to deceive the other person and in turn trying to guess when the other person was lying to them; they sat in front of each other at a table with a screen on the table so that only the top half of their bodies and their heads were visible to the other; senders were aware of being recorded – each of them was recorded by a camera that was placed at the opposite person's side, therefore the videos were at an angle and it was normal that senders were not looking directly at the camera as they were conversing with the other person.

Three video tasks were created from videos recorded in Stage 1. The Demeanour Task (DT-B) was the first task receivers attempted, followed by a quick 5-minute break and then the Weird Pictures – Baseline (WP-B). There was a ten-minute break after the WP-B. The session resumed with two blocks of Two Truths and a Lie – Baseline (TTL-B) with a 5-minute break in the middle.

At the end of the of the testing session, participants were given the option to photograph their response sheets and collect an answer keys once they turned in their response sheets – this gave them the opportunity to find out their scores on the deception-detection tasks, if they so wished. They were debriefed about the general purpose of the study and were individually remunerated for their time, if they had signed up for monetary reward.

2.2.2.2.1. Weird Pictures – Baseline (WP-B)

2.2.2.2.1.1. Task creation

The Weird Pictures task for Stage 2 (WP-B) was created by making sets of three videos (triplets) for each sender who took part in Stage 1 – one Truth, one Lie, and one Double Bluff (DB). The Lie and DB descriptions were taken from WP-P, whereas the Truth descriptions were extracted from WP-Control.

Strict shortlisting criteria controlled the quality and appropriateness of videos. Any videos that met the following criteria were excluded:

- Technical issues – blurred video, excessive background noise, audio indecipherable.
- Where the sender did not follow instructions and did not adhere to timings marked by bell: where they either started before the bell or continued speaking after the bell or where the experimenter can be heard prompting them because they failed to start at the correct time.
- Where the description a senders claimed to be a lie was not sufficiently different from the actual content of the picture they were describing. This could include changing only one or two elements of the picture or changing relatively minor elements, such that the picture remained recognizable from the content of their description. Alternatively, where senders claimed to have double-bluffed, however their descriptions deviated from the actual picture in some way.
- For WP-Control, videos where the sender had directed receivers (from Stage 1) to look for certain elements of the picture (in order to recognize the picture from the grid of nine pictures they were presented) were excluded.

- As the videos were to be part of deception-detection battery for participants under the age of 18 years, any videos that violated the BBFC's Universal Rating (*BBFC | Ratings at a Glance*, n.d.), such as, any mention of alcohol, smoking, sex, swearing etc., were excluded.

- Since senders had a choice about how they wanted to deceive in WP-P, some participants chose to lie on all five trials or double bluff on all five trials. These participants were excluded since a viable triplet of Truth-Lie-DB could not be formed for them. Similarly, any participants who did not have any usable Truth, Lie, DB (due to any of the above-mentioned criteria), were also excluded.

If a sender did not use the full 30 seconds to describe the picture, but that were descriptive enough (regardless of whether they lied / double bluffed / told the truth), the video was not excluded. Instead, the video was edited such that the silence at the end of the clip was removed and only the portion where the sender was speaking was included.

After all exclusions, videos were arranged in triplets. Some senders had more than one viable triplet (n = 9), while some had only one viable triplet (n = 14). A total of 32 usable triplets were created, divided equally amongst 4 batches. Batches were created keeping in mind that every image was only described once across all senders in that batch, i.e. all DBs and Truths were descriptions of images that only featured once in that batch. This was done to avoid any repetition of the picture content and consequently helping receivers gauge the authenticity of the descriptions.

Each batch had eight trials, i.e. one unique triplet from eight different senders. The videos (Truth-Lie-DB) within each triplet were presented in the same pseudorandomized order, as were the eight triplets in each batch (Figure 2.4).

Figure 2.4

Graphic illustrating the trial and batch set-up for WP-B

Note. Each box represents a unique video that appears only once across all trials and all batches. DB = Double Bluff, L = Lie, T= Truth

2.2.2.2.1.2. Task administration

Receivers in each batch were told that they would be shown videos of eight people, each describing three pictures. An example of one such "weird" picture was shown. They were informed that one of the three descriptions each person gave would be a true description, one would be a lie, and one would be a double bluff (i.e. the person would describe the picture truthfully but pretend as if they were lying); the three descriptions could be in any order; and their aim as receivers was to guess which description was the lie.

Each video in the triplet was preceded by a number – '1', '2', or '3' (in that order). Participants had 10 seconds after each set of videos to log their responses on the response sheet by circling the number corresponding to the video they thought was a lie. At the end of the 10 seconds, a chime would sound and the next set of videos would start playing. Finally, receivers were shown a sample triplet to help them understand the task better, and were asked if they had any questions.

The WP-B took roughly 15 minutes to complete.

2.2.2.2.2. Two Truths and a Lie – Baseline (TTL-B)

2.2.2.2.2.1. Task creation

The Two Truths and a Lie task for Stage 2 (TTL-B) was created from the TTL-P responses by senders in Stage 1. Since every sender gave three statements (two true statements and one lie) on each of the four topics they were given, the entire trial, i.e. three statements (triplet), was kept intact and the statements remained in the same order that they were delivered by the senders.

Similar to WP-B, a strict shortlisting criteria controlled the quality and appropriateness of videos. Any videos that met the following criteria were excluded:

- Technical issues – blurred video, excessive background noise, audio indecipherable.
- Where the other participant (with whom the sender was playing TTL-P) is heard making remarks other than a simple acknowledgement (e.g. "yeah"), where the sender refers to and piggybacks off something the other participant had said in a prior trial, or where there is a conversation between the two participants.
- Where the sender did not follow instructions – did not adhere to the timings marked by the bell, where they either started before the bell or continued speaking after the bell or where the experimenter can be heard prompting because they failed to start at the correct time.

- As the videos were to be part of deception-detection battery for participants under the age of 18 years, any videos that violated BBC's Universal Rating, such as, any mention of alcohol, smoking, sex, swearing etc., were excluded. In addition, any videos with controversial or offending material, such as references to politics (such as particular politicians and parties) or religion (in a derogatory context or otherwise), were excluded.

Where even one of the three statements met one of the exclusion criteria above, the entire triplet was discarded. Triplets were also excluded if the senders confessed after the trial or during the debriefing session to have mixed up / mislabelled the true and lie statements.

After all exclusions, four senders had only one viable triplet, 11 senders had two viable triplets, 18 senders had three viable triplet, and two senders had all four triplets that were viable. A total of 88 usable triplets were shortlisted, divided equally amongst four batches, i.e. 22 triplets per batch.

Batches were created keeping in mind that no two TTL-B triplets were from the same sender. Only two (out of 22) triplets in batch featured senders who also featured in a WP-B triplet in the same batch. The number of triplets related to each of the four topics – Weekend, Future, Hobbies, and Summer – were consistent across all batches, give or take one.

To avoid fatigue, the task was split in two blocks each consisting of 11 trials. The 11 triplets in each batch were presented in the same pseudorandomized order for all participants in that batch.

2.2.2.2.2.2. Task administration

Receivers in each batch were told that they would watch videos of the senders playing the icebreaker game "Two Truths and a Lie", where they answered a question with two true answers and one lie in no particular order. Their aim as receivers was to guess which answer was the lie. As with WP-B, they had 10 seconds to jot down their responses on the response sheet at the end of which a chime would sound, indicating the start of the next trial. They were then shown a sample triplet.

Receivers were informed that they would be watching videos from a total of 22 people, but, in order to avoid fatigue, the task was broken into two blocks of 11 triplets between which they could break for five minutes.

Each block of TTL-B took about 20 minutes to complete.

2.2.2.2.3. Demeanour Task – Baseline (DT-B)

The Demeanour Task – Baseline was created as a measure of the demeanour of the senders who featured in WP-B and TTL-B, i.e., how honest or dishonest they looked in general when not actively deceiving.

For videos to be appropriate to gauge demeanour, senders must be judged regardless of whether they were lying or not (T. R. Levine, 2010). Therefore, video snippets were selected from Stage 1 where the senders were not engaged in deception tasks, i.e. from the segment where they were asked debrief questions following WP-P and TTL-P. Considering that senders spoke about lie-production and deception-detection strategies during the debrief sessions, the videos prepared for DT-B were silenced (i.e. the audio was removed) to avoid influencing judgement based on the content of the discussion.

Across all four batches of both WP-B and TTL-B, there was a total of 35 unique senders. Therefore, 35 silent videos were created, each 20-seconds long. The DT-B in each batch consisted of videos from 28 senders - eight of who featured in WP-B and 20 additional ones from TTL-B (two senders in TTL-B were also a part of WP- B; true for all four batches).

Since demeanour-judgement must be uninfluenced by deception ability, DT-B was the first video task in Stage 2. Receivers were told that they would watch silent videos of 28 people, who were having a conversation with another person off camera, and they were required to rate how honest or dishonest they thought senders were. Responses were recorded on a response form where each sender was rated on a Likert-type scale from 1 (most dishonest) to 7 (most honest). As with all other video tasks, receivers were given 10 seconds to respond and a chime indicated the end of response time and start of next trial. Receivers were shown a sample DT-B video to familiarise them with the task.

The DT-B took roughly 15 minutes to complete.

2.2.3. Stage 3: Pilot studies

The pilot studies detail the process of how the data from the baseline study was used to create online version of the deception detection tasks, to be administered in the future developmental and autism studies.

2.2.3.1. Pilot-1

2.2.3.1.1. Participants

Sixty-four participants (M_{age} = 18.88 years, SD_{age} = 0.97 years, 55 Females, 9 Males), recruited through the UCL Psychology and Language Sciences online participant database, took part in the first pilot study. Participants were awarded course credits for their participation.

2.2.3.1.2. Materials and Procedure

The performance of the receivers in the Baseline tasks, i.e. WP-B and TTL-B in Stage 2 were used to shortlist triplets for the deception detection tasks that would be administered online in Stage 3.

For each sender, receiver success on each of their triplets was averaged (weighted average taking into consideration that different batches in the Baseline study had different number of receivers). Receivers were scored 0 when they failed to detect lie in a triplet and 1 when they detected the lie successfully.

To mimic Levine et al. (T. R. Levine et al., 2011), I attempted to create two sets of videos: transparent – i.e. senders that were easiest to judge – and non-transparent – i.e. senders that were hardest to judge. Considering demeanour was constant within each triplet, it was not possible to create honesty-veracity matched and mismatched conditions, i.e. transparent (honest-truthtellers and dishonest-liars) and non-transparent (dishonest-truthteller and honest-liars) conditions akin to tasks created by Levine at al. (2011). Instead, the transparent and non-transparent senders were selected based on how easy or how difficult receivers across all batches from Stage 2 – Baseline study found the senders. Four senders whose lies were judged correctly most often by receivers in Stage 2 were selected as transparent senders. Similarly, four of the most difficult to judge senders were selected as non-transparent senders. it was assumed that the senders hardest to judge leaked the least amount of cues and thus were non-transparent and that the senders the receivers had most success with were transparent as they leaked more cues. One video triplet for each of these eight senders were then randomly selected such that the final stimuli set had 4 easy-triplets, i.e. the triplets from transparent senders, and 4 hard-triplets, i.e. triplets from non-transparent senders. Furthermore, considering that the paradigm being developed was to be administered subsequently to a developmental sample, non-native English speaking senders were not shortlisted as part of the paradigm in order to remove the added factor of sender fluency and accent, especially as there is some indication that people take longer to evaluate statements by non-native English speakers (Munro & Derwing, 1995).

The pilot test session was created and administered via Gorilla Experiment Builder (www.Gorilla.sc). Participants completed the tasks online on a computer at home.

Both the WP and the TTL had eight triplets and the videos within each triplet were presented in the same fixed order as they were in the baseline study. The eight triplets were presented in random order. Participants were given 5 seconds after watching the triplet to respond by clicking one of the three buttons (labelled 1, 2 and 3 indicating the three videos in the order they appeared). Participants were also shown a sample triplet at the beginning of each task to help familiarise them with the tasks. The deception tasks were part of battery of other tasks and questionnaires, and the whole session took about an hour to complete.

All procedures were approved by UCL Research Ethics Committee (Project ID Number: 14807/002).

2.2.3.1.3. Analysis and evaluation of Pilot-1

An above chance performance (chance = 0.33 x 4 = 1.33 i.e. one in three chance of getting it right multiplied by four trials) from the non-autistic adults sample was indicate that the task was working as expected.

This first pilot study showed that participants performed at chance on the WP: t (63) = -1.306, p = .196, and significantly below chance on the TTL: t (63) = -2.583, p = .012. When analysing each item, floor effects were observed for all triplets in the WP – i.e. receivers performed at chance (= 0.33) or below chance level. Only performance on one hard-triplet showed above chance performance; one-sample Wilcoxon signed rank test: $z = 649$, $p = .01$. Similarly, performance on the TTL was also at or below chance for all but one hard-triplet; one-sample Wilcoxon signed rank test: $z = 754$, $p = .041$.

Therefore, it was concluded that the tasks in their present condition were not successful. Feedback from participants and re-evaluation of the task designs in light of the findings, led to the following conclusions:

- The tasks were too long at roughly 20-25 minutes each. Participants felt their concentration waning. Since tasks in the Pilots were online unlike Stage 2 where they were offline, attention spans of participants were likely to be even shorter.
- One possible reason that the tasks in Pilot-1 did not work could be that Pilot-1 did not include a demeanour task, unlike Stage 2 – Baseline study. The demeanour task (DT-B) administered before the WP-B and TTL-B in Stage 2 may have had a spill over effect of familiarity, thus affecting receivers' performance. It is well documented that prior exposure to non-deceptive behaviour of senders increases deception detection ability of receivers (Brandt et al., 1980, 1982; Feeley et al., 1995).
- Receivers had a tendency to select the second (middle) video in triplets as their response. In both tasks, the only triplets to have above chance performance had the lie placed in second position in the triplet. This could be a reason why even though baseline receivers performed poorly on triplets from these senders, the receivers in this pilot were most successful on these trials.

2.2.3.2. Pilot-2

2.2.3.2.1. Participants

Ninety participants (M_{age} = 26.98 years, SD_{age} = 9.58 years, 52 Females, 28 Males) were recruited either through the UCL Psychology and Language Sciences online participant database or through Prolific (www.prolific.co). Participants were either awarded course credits or the departmental rate of £7.50 per hour. All participants were non-autistic.

2.2.3.2.2. Materials and Procedure

For the second pilot, triplets from the Stage 2 - Baseline study were shortlisted with fresh set of criteria: instead of averaging receiver success from Stage 2 across all triplets from each sender, each triplet was considered as an individual item. Only four triplets for each task were selected to shorten the length of each task. The four triplets that the baseline receivers had the most success on were selected to avoid floor effects. Ceiling effects were unlikely, as they were not observed for these triplets in the baseline study.

Two versions of the WP and TTL were created – one where the videos within a triplet were randomised between participants (Random-order) and one where the videos were presented in same order as they were in Stage 2 – Baseline study (Fixed-order). Both versions used the same triplets, only the order of the videos within the triplets was different between the two versions. This was done to control for participants' tendency in Pilot-1 to respond by selecting the middle video more frequently than the other two, thereby affecting the detectability of individual items. Participants were randomly allocated to either versions of both TTL and WP so that some participants took part in the Random-order TTL and Fixed-order WP, and vice versa, and some took part in the same order condition for both tasks.

Both versions of WP and TTL had four trials each; every trial consisted of one triplet. Participants were given 5 seconds after watching the triplet to respond by clicking one of the three buttons (labelled 1, 2 and 3 indicating the three videos). Participants were also shown a sample triplet at the beginning of each task to help familiarise them with the tasks. The deception tasks in the second pilot took 10-12 minutes each.

A demeanour task, consisting only of the senders appearing in the shortlisted WP or TTL triplets, was administered before the deception tasks so that any possible effects of familiarity that may have affected receiver judgements in Stage 2 – Baseline study would also play contribute to receivers performance in Pilot-2. This task took about 4 minutes to complete.

The pilot test session was created and administered via Gorilla Experiment Builder (www.Gorilla.sc). The deception tasks were part of battery of other tasks and questionnaires, and the whole session took about 40 minutes to complete. Participants completed the tasks online on a computer at home. All procedures were approved by UCL Research Ethics Committee (Project ID Number: 14807/002).

2.2.3.2.3. Analysis and evaluation of Pilot-2

Due to the fewer number of trials in Pilot-2, the data was considered ordinal and non-parametric tests were conducted instead of the standard one-sample t-tests. Four one-sample Wilcoxon Signed Rank tests were performed to see if participants performed above chance on WP and TTL in the Random-order and Fixed-order versions of the task. Considering we only used easy-triplets and participants were non-autistic individuals like in the baseline study, it was expected that for the tasks to be deemed working participants would need to perform at a level significantly above chance on the overall task (chance = 0.33 x 4 = 1.33 i.e. one in three chance of getting it right multiplied by four trials). Performance was significantly above chance on all tasks (Table 2.2).

Table 2.2

One-sample tests to investigate if performance was above chance on deception tasks

Task	Type	*n*	*M* (*SD*)	*Mdn*	*T*	*z*	*p*	*r*
WP	Random order	44	1.75 (1.08)	2	728.00	2.74	.006	0.41
	Fixed order	46	1.67 (0.85)	2	826.00	3.17	.002	0.46
TTL	Random order	45	2.33 (1.02)	2	965.00	5.10	<.001	0.76
	Fixed order	45	1.69 (1.13)	1	705.00	2.14	.033	0.31

Note. One-sample Wilcoxon Signed Rank tests against test-value of 1.33 (chance: 4 trials x 1/3 probability of getting the correct answer by chance).
n = sample size *M* = mean, *SD* = standard deviation, *Mdn* = Median, *T* = Test-statistic, *z* = z-score, *p* = *p*-value, *r* = effect size.

In both the WP and the TTL, participants performed above chance for both the random- and fixed-order versions. Therefore, it can be assumed that some combination of factors – using easy-triplets, adding a demeanour task that increased familiarity, and shortening the task improved performance on the task.

While for the WP both the random- and fixed-order versions have similar effects sizes, the random-order TTL had a large effect size compared to the small effect size of the fixed-order TTL. Therefore, the decision to retain the random-order versions of WP and TTL was made.

The final versions of the deception detection tasks created – WP-Final and TTL-Final – have four trials each, trials are randomised between participants and each trial consists of a triplet of videos that are presented in a random order. The tasks each take about 10-12 minutes.

2.3. Discussion

Studying deception in a controlled way in a laboratory setting is a difficult feat, not only considering the ethical issues around it (such as informed consent), but also because of the myriad of ways the validity and generalisability of deception paradigms are challenged. I created two deception production tasks that attempted to address the shortcomings of past deception studies and further created two novel deception detection tasks using the stimuli from the tasks in the production study. The tasks were designed to tackle issues that previous laboratory based deception studies have faced such as lack of motivation, choice, and interaction. However, the novel tasks designed here also had some areas that can be further improved on. In this section, I discuss the various characteristics of the two tasks and reflect on how they strengthen the experimental design. I also consider some of the shortcoming of these tasks and comment on how I have tried to mitigate them and how they can be improved upon in future iterations.

2.3.1. Ground Truth

To be able to measure correctly either the senders' capability to deceive or the receivers' capability to detect that deception, experimenters have to be confident in the distinction between deceptive and truthful stimuli. They need to be aware of the *ground truth* behind the lies that the senders tell, i.e. the reality of what actually happened, what the sender actually feels or what they actually think, what they are actually supposed to be describing etc. Unlike studies where senders are asked to lie about their opinions on a given topic or give a false description of a person they dislike, describing a picture such as in the Weird Pictures Task ensures that we can be confident about the ground truth (Frank, 2008). Additionally, we were able crosscheck and discard any instances where a sender deviated from the ground truth in either the truth condition (WP-Control) or the double bluff condition (WP-P) when shortlisting stimuli for the WP-B.

Contrarily, the ground truth in TTL-B could not be confirmed and the experimenters had to rely on the senders to confirm that their truth statements were actually true. This is problematic because senders could pass off statements that were not completely true to portray themselves in a better light (concealing unflattering elements or exaggerating elements to make themselves look better), or could be untruthful about which of their statements was the actual lie (intentionally mislabel the lie) to win the bonus monetary reward for that round. One way to eliminate the possibility of *intentional* mislabelling in the future would be to have the sender record (electronically or on paper) which statement was a lie *before* the receiver gives their response.

We attempted to mitigate any *unintentional* mislabelling by asking participants during the debriefing session that their truths were actually true. One participant admitted to having either lied twice in a trial (instead of telling two truths and one lie) and one mentioned not being able to come up with a lie that was more than just an omission of the truth – both these trials were excluded from TTL-B. It can be presumed therefore that participants who did mislabel their truths for lies would have admitted doing so during the debriefing session.

2.3.2. Double Bluff

Lying is cognitively taxing (Vrij & Ganis, 2014) – senders must inhibit the urge to tell the truth (Debey et al., 2015) and must hold in working memory the content of the truth and the lie they want to say (Vartanian et al., 2012). Interrogators use a number of techniques based in increasing the cognitive load to detect lies better (Vrij et al., 2017). Therefore, it is understandable that senders would opt for the cognitively easiest option that fulfils the situational demand.

The option to double bluff in the WP-P offered senders a less cognitively taxing option to meet the brief of the game, which was to deceive the receiver. By choosing to double bluff, they did not have to inhibit the truth, as they would have to if they chose to lie, and only had to hold the truth in their working memory instead of both the truth and the planned lie. The fact that some of the senders still chose to lie instead of double bluff shows serious intent and commitment to lie. Therefore, it can be assumed that on trials that senders lied in the WP-P, senders were motivated and confident in their lies, unlike in traditional deception experiments where senders are not invested in their lies. Alternatively, this could have also been a part of he senders overall strategy – to intermix lies and double bluffs to be unpredictable, which too shows that the participants were fully invested in the game. Second, since there were some senders who chose not to lie at all during all the five trials of the WP-P, it leads us to believe that the option of double-bluffing effectively eliminated any senders who were inherently under-confident in their lying abilities and might choose not to lie in real life as a rule of thumb. Therefore, while creating our deception detection task, WP-B, we did not have any sender who was only lying under the pressure of the task instructions.

As part of deception detection task, a double bluff condition provides a good control for content-driven lie detection in the WP-B. When it comes to the stimuli used, one known factor that drives deception detection is the content of the lies. This is evident in the fact that receiver’s detection accuracy improves when lies are contextually relevant (Blair et al., 2010) and that detection paradigms that allow for cross-questioning / follow up questions, thus increasing the contextual content to be judged, show better accuracy (T. R. Levine et al., 2014). This is also evident in differential rates of success in deception detection when receivers rely only on content based cues; receivers’ detection rates fall when senders have time to plan their deception except when receivers rely only on content cues (Littlepage et al., 1986), thus showing that content cues help receivers mitigate the added difficulty due to preplanning of lies. The double bluff statements were true descriptions of the pictures, therefore ensuring that their content was similar (except for individual sender differences) to the truth condition stimuli, therefore proving to be a good content-based control in the WP- B.

It can be argued that the truth statements in TTL-P were also double bluffs. Considering that the senders were providing the truths and the lie with the goal of making it hard for the receiver to pick out the lie, it can be assumed that they modified their mannerisms to conceal that the true statements were actually true, thus instead double bluffing. Since the receivers were aware of the true nature of the game, and their goal was to guess which one of the three statement is the lie rather than making truth-lie judgements on individual statements, it could be assumed that it did not affect their performance on the TTL deception-detection tasks.

2.3.3. Self-selection and choice

Allowing participants to self-select in deception tasks, especially in a study aimed at stimuli development, might lead to a loss of experimental control, especially when they may choose to tell the truth instead of deceiving as they lack the motive to deceive in the first place. On the other hand, randomly instructing senders to deceive by assigning lie/truth conditions to participants may change the behaviour of the senders enough that it no longer reflects non-experimental real-world deception behaviour (T. R. Levine, 2018). The deception games created for this project attempted to balance the two. The games ensured a certain level of experimental control by putting the senders in a situation where deception is the goal, thus, regardless of what choices they made they were deceiving on every trial. Further, success had real consequences (bonus monetary reward) and this guaranteed that senders were actively engaged in deceptive behaviour by giving them a motive to deceive successfully. Additionally, it also mitigated the behavioural change that might have come with instructing (and therefore compelling) senders to deceive by embedding choices about how to deceive within the games so that senders may self-select.

As mentioned above, introducing the option to double bluff in WP-P allowed participants to choose the method of deception, and therefore, when they did lie, it was self-selected. In TTL-P, participants strategized about the sequence of their description and chose when to lie. Not only does giving participants a choice about order gives them more control and thus increases motivation (Patall et al., 2008), it also increases the ecological validity of the task. According to Frank (2008, p. 16), “lies and truths are often intertwined” and their placement in respect to each other is not defined by external sources, but by the sender themselves. Therefore giving the senders in our task the flexibility to choose *when* they wanted to lie allowed us to collect stimuli that was more ecologically valid than when orders were predefined.

While self-selection helps increase ecological validity and motivation, it does (as opposed to random assignment) bring in to question the internal validity of a study by introducing external factors such as the senders mood, personality etc. (Dunbar et al., 2015). However, this is mostly an issue when veracity is a between subjects factor – like most traditional deception experiments. As our tasks do not have senders self-select into either a truth or a deception condition but contrarily include deceptive and non-deceptive stimuli from all senders (WP-B and TTL-B), this is unlikely to be an issue as any external factors will be common to deceptive and non-deceptive stimuli alike.

2.3.4. Sanctioned lies

We told the participants in the lie production study (stage 1) that they would be playing some ‘deception games’ with the other person. This was done in order to encourage interaction, completion, and motivation. A complication that ‘gamifying’ deceptive behaviour presents is that it introduces an explicit permission to lie, unlike most non-experimental real-life situations where people deceive. These are called sanctioned lies – deceptive episodes where the sender has been instructed to lie and they differ from unsanctioned lies – where people are not authorised to lie but decide to deceive because they believe deception to be the mode of communication that aligns the most with their communication goals.

When people decide to deceive in real-life scenarios, it is usually unsanctioned. Miller and Stiff (1993) propose that unsanctioned liars would show higher motivation versus sanctioned liars as unsanctioned liars may experience negative feelings of guilt and shame whereas sanctioned liars do not feel the same moralistic repercussions and externalise the blame for their deception on the researchers rather than internalising it. Feeley and deTurck (1998) suggest that sanctioning lies removes elements which are crucial to unsanctioned lies – accountability, fear of consequence, motivation – which all lead to heightened arousal and a bigger cognitive load.

As Sporer and Schwandt (2007) note in their meta-analysis, there are very few studies investigating if the sanctioned or unsanctioned nature of deception has any effect on deceptive behaviour and its detection. However, they did find that unsanctioned liars smiled more than sanctioned liars, although they caution that the results are derived based on a very limited number of studies investigating unsanctioned lying. It has also been found that deceptive behaviours such as speech hesitations and errors, and gaze differed when sender told unsanctioned and sanctioned lies about having cheated on a quiz (Feeley & deTurck, 1998).

More recently though, Dunbar et al. (2015) directly compared sanctioned and unsanctioned deception in a high-stakes paradigm and found that there was no difference in how credible or expressive senders appeared in the two conditions. They also found that expert receivers –

certified polygraph interviewers – did not find lies of one group more difficult to judge than the other group. Additionally, sanctioned and unsanctioned liars did not differ in how guilty they felt about their deception, contradicting Miller and Stiff's (1993) assertions that guilt is one of the drivers of any difference between sanctioned and unsanctioned lying.

Therefore, the gamified nature of the novel deception tasks created here might not necessarily affect their ecological validity as the senders in our sanctioned deception games may not have behaved any differently than senders in real life partaking in unsanctioned deceptive behaviour. If expert senders in Dunbar et al.'s (2015) study did not find any difference between sanctioned and unsanctioned liars – except when they asked them follow up questions – it is highly unlikely that the deception detection abilities of the non-expert receivers in any of our detection studies would have been affected by the sanctioned nature of the deceptive stimuli they watched.

2.3.5. Generalisability of content

Generalisability of findings from a lab-based study to non-lab based ecology is always a concern in any psychology research as controlled laboratory conditions and experimenter effects may change behaviour in a lab from that outside. There are, however, various ways in which these could be mitigated, as demonstrated above, all the while affording the control that lab-based deception studies provide over naturalistic studies.

This is also true for choosing what the senders would be lying about. In the WP-P task, senders are required to deceive about a picture in each trail. This not particularly representative of deceptive behaviour outside of the lab (Frank, 2008). This also means they have no personal attachment with the picture or emotional investment in their deception. This is partially moderated by the introduction of choice and the chance to win extra monetary incentives, however this must be taken into consideration when generalising findings to real life. That said, it is important to remember that this loss of ecological validity comes with enhanced experimental control – it allows us to have full confidence in the ground truth and control for content-driven deception detection with the help of the double bluff condition.

The TTL-P, in this regard, is more ecologically valid, as senders are more likely to deceive about their future plans or hobbies to a stranger than about pictures. While still within a sanctioned environment, divulging personal autobiographical information likely means that they are more aroused and emotionally invested in the lies and therefore are in an emotional state more akin to real-life deception (versus the WP- P).

2.3.6. Receiver Fatigue

The baseline study was integral to both the deception production study to ensure that each sender's production capability was not dependent on the receiver who took part with them. It was also an essential step towards shortlisting stimuli for a deception detection paradigm that could be implemented on a wider and varied sample. It was rather designed such that one group of receivers did not have to judge stimuli from all the senders in stage 1, which would have amounted to a testing session with a run time of over 7 hours. Instead, receivers participated in batches and watched stimuli that were unique to their batch. This ensured that all the stimuli were judged by a group of receivers to increase objectivity and reduced the length of the test session. Even so, the baseline study for each participant was two hours long. According to Frank (2008), deception detection for any length of period over 30 minutes could induce fatigue in receivers. We attempted to alleviate this by breaking longer video tasks into smaller blocks and giving frequent breaks in between the different tasks. While within a batch all participants watched the videos in the same order, any senders who had triplets in more than one batch appeared in different place in the order in every batch. It was hoped that averaging deception scores across batches for each sender would counteract the detrimental effect that receiver fatigue may have had; however considering that the final shortlisting of WP and TTL (Pilot-2) did not shortlist triplets by averaging scores across batches, receiver fatigue remains a concern.

In future, a baseline study of this nature could be done with more batches, to spread out the stimuli and reduce the testing time for every batch. Alternatively, instead of running the baseline study in batches, receivers could be tested individually, in a between-subjects design. These methods however could increase the running time and costs and therefore might not be feasible for large stimuli sets such as this.

Receiver fatigue was also a likely concern for Pilot-1. Feedback from participants highlighted that they found the tasks too long and their concentration waned during the task. To counter this, Pilot-2 was shorter – only four trials in each task instead of eight – and this could be one of the factors that contributed to the above chance performance in the second pilot.

2.3.7. Effectiveness of the tasks

Meta-analysis of deception research shows that deception detection rates are only at about 54% across studies, and that while this is above chance, the effect sizes are usually small (C. F. Bond & DePaulo, 2006). From Pilot-2, it is evident that receivers are significantly above chance on the novel tasks created here – with medium to large effect sizes. One reason for this could be that the final versions of the tasks, i.e. the ones piloted in Pilot-2 were made of the triplets that receivers in Stage-2 found easiest to judge. Levine (2010) claims that the reason why people are over chance at detecting deception is because of the presence of transparent – i.e. easy to detect – liars amongst non-transparent liars. Given that the WP-Final and TTL-Final are made up of only transparent senders, it is understandable why the detection rates are higher than what is found in previous literature.

A possible argument against the effectiveness of the task therefore could be that it deviates from how individuals experience deception in daily life – as a mix of transparent and non-transparent receivers. However, considering that people are below chance at detecting deception from non-transparent senders (e.g.: D. M. Williams et al., 2018), these tasks provide an opportunity to capture any variance in deception *detection* abilities without encountering floor effects. This, along with the fact that these task control for sender demeanour, which has been as one of the strongest predictors of variance in deception judgement (C. F. Bond & DePaulo, 2006), these tasks show potential as a deception detection task that measures individual differences in receivers rather than senders.

2.4. Conclusion

In this chapter, I showcased the various problems with past deception studies and the paradigms they use and highlighted the need to create new tasks to produce deceptive behaviour, generate better stimuli and design different deception detection paradigms. I created two deception tasks – the 'Weird Pictures' (WP) and the 'Two Truths and a Lie' (TTL). They both had a version that was used to produce deception (WP-P and TTL-P), that contributed to the creation of two novel deception detection tasks – WP-Final and TTL-Final – that attempt to correct some issues with past deception detection tasks.

Each of the two tasks has certain elements that make them more internally and ecologically valid than previous lab based deception detection experiments, as discussed. The two task do not claim to exactly mimic deceptive behaviour (and therefore also detection of deception) in the real world, nor do they individually manage to mitigate every single pitfall of previous deception tasks. In my opinion however, as a pair, the tasks complement each other in strengths and shortcomings and I felt that using them together would help develop an accurate profile of receivers' deception-detection abilities

Chapter 3. Deception Production

3.1. Introduction

Deception is a communication tool that everyone has in their arsenal; however, people do not deceive indiscriminately. There is strategic decision making involved regarding when one uses deception – people only deceive when it is the option that provides the least amount of resistance towards fulfilling their goals (McCornack, 1997). It is used when the truth will not suffice or is not efficient, and often as a last resort (T. R. Levine, Kim, et al., 2010). But, once a person does decide to deceive, what are the factors that help them decide how to deceive and what are the factors that govern whether or not they are successful in their deception? Likewise, are there some inherent properties that make some people worse at deception?

There are various ways to deceive, for example, omission, part lies, falsification, double bluffs etc. Furthermore, different types of deception require different sets of or different levels of cognitive processing (Morgan et al., 2009). Similarly, the subject of the deception may affect cognitive processing (Ofen et al., 2017) and therefore people might vary in how well they lie when lying about different topics. Therefore, in this chapter, I will use the novel tasks I created in Chapter 2 to explore if factors affecting deception ability, such as cognitive load, also affect people's choices about how to deceive. I will also investigate if having to deceive about impersonal or personal subjects affects peoples' skill at deception.

In Chapter 1, I introduced the idea that certain cognitive abilities, such as how well senders understand others' mental states (i.e., mentalizing), and their individual characteristics, such as how honest or dishonest they appear (i.e., demeanour), are thought to affect deception success. I plan to examine these claims in this chapter by directly testing the relationship between mentalizing and deception, as this has rarely been done in adults before. I will also explore if demeanour predicts deception success, like Levine (2010) insists, even when demeanour is kept constant across the different veracity types.

As demeanour, i.e., how honest or dishonest a person is perceived to be, is presumed to play a role in deception success, it is possible that autistic traits also influence deception success, considering autistic individuals are generally perceived more negatively than non-autistic individuals (e.g.: Alkhaldi et al., 2021) and might therefore be judged to have dishonest demeanours regardless of their actual propensity to lie. Autistic individuals are also thought to have difficulties in mentalizing and producing deception (e.g.: Sodian et al., 1991; Sodian & Frith, 1992), although their deception production success has rarely been explored. My final aim in this chapter, therefore, is to investigate if autistic traits are related to how dishonest a person is judged to be, regardless of their actual truthfulness, and if self-reported autistic traits are related to deception success.

3.1.1. Effects of cognition on deception production

Deception is cognitively effortful (Vrij & Ganis, 2014). Pupil dilation, associated with increased cognitive effort, is greater when senders lie versus when they tell the truth across different types of lies (Dionisio et al., 2001). Cognitive cost, as operationalised by reaction times across various paradigms, has also been seen to be higher in conditions where senders are required to lie as compared to when they are truthful (Suchotzki et al., 2017). Deception is also linked with increased activation in the executive regions of the brain, primarily in the prefrontal and anterior cingulate cortices (Zeki et al., 2004). Furthermore, as covered in Chapter 1, various different techniques of increasing deception detection rates involve increasing the cognitive load on the sender (e.g.: Vrij et al., 2008, 2010) so as to force the sender to leak more cues due to the extra mental effort. So, it can be extrapolated that cognitive effort will also affect other aspects of deception, beyond just ability.

3.1.1.1. Type of deception

Deception may be cognitively costly, but are the cognitive costs the same even when one is not overtly lying? Double bluffing is a way to deceive where the message delivered by the sender is truthful in content but is delivered with the intension to mislead the receiver into thinking that the contents are untruthful. In other words, a sender double bluffing would tell the truth but would manipulate their mannerisms to mimic those they would display if they were lying.

Most research related to double bluffing has been in the context of theory of mind development - double bluff is used as a measure to test individuals' understanding of higher order mental states, often as part of mentalizing batteries (e.g.: S. White et al., 2009). A handful of studies, however, have investigated double bluff alongside lying and other forms of deception. Neuroimaging studies have found that honest manipulation (viz. double bluff) has a different pattern of brain activity to lying (Ding et al., 2014; Kireev et al., 2017), although other fMRI studies have demonstrated that deceptive intent – regardless of whether it is by lying or using honest manipulation – triggers similar neural activity (Volz et al., 2015; Zheltyakova et al., 2020). Sip et al. (2010) suggest that while there are common neural patterns in both types of deception, false claims (vs true claim, similar to double bluff) put additional decision making load on the sender that manifests as greater neural activity in regions associated with selection and execution of actions, such as the premotor and parietal cortices. Contrarily, Sutter (2009) proposes that a complex decision making process based on what the receivers expects is present even when senders choose to tell the truth – either as a deception tactic or genuinely. Therefore, these mixed finding demonstrate that our understanding of the commonality between lying and double bluff when it comes to neurocognitive processes is limited.

While there has been some investigation of the neural correlates of lying versus double bluffing, no study has directly studied if there is a difference in the cognitive effort involved in double bluff in comparison to lying that manifests as behavioural differences between the two, such as having a bias of choosing either double bluff over lying. Lying, where one provides false information (falsification), forces the sender to utilize inhibitory control in order to stop themselves from delivering the truthful message (Debey et al., 2015) and tax their working memory in order to remember the truth and also the alternative untruthful message (Vartanian et al., 2012). Double bluffing, on the other hand, theoretically, does not require as much cognitive effort as lying considering senders only need to hold the truthful information in their working memory and do not need to inhibit themselves from delivering the truth. Therefore, it is easy to extrapolate that if given a choice, people might choose the less taxing way to deceive – to double bluff instead of lying outright.

This study aimed to directly tests this in a paradigm where senders had a free choice of how they wanted to deceive about pictures they had to describe to another person. Based on the idea that double bluffing is possibly less cognitively effortful than lying, it was predicted that,

> **H_1:** Regardless of deception success, participants would choose to double bluff more often than to lie. In other words, when given a choice between lying and double bluffing, such as in the Weird Pictures Task (WP), senders would have a bias towards choosing to double bluff.

3.1.1.2. Content of deception

Frank and Ekman (1997) proposed that cognitive effort becomes more visible, i.e. leakage of cues becomes more prominent, in deceptions where the stakes are high and the sender has a lot to lose. It is generally agreed upon that high stake lies cause senders to leak more cues versus instances where the stakes are lower, thus making senders in high stake scenarios relatively easy to detect. This is thought to be because the higher the stake or emotional attachment, the more the cognitive effort of controlling emotional cues of deception, and consequently higher the risks of leakage (Frank & Ekman, 1997).

The variability in deception success has been investigated before in terms of the stakes and severity, such as the study by Fuller et al. (2015) where they found that, as the severity of the lies increase, senders leak more cues. The variance in deception success as a function of content of lies on the other hand has rarely been tested directly. .Nuñez et al. (2005), through brain imaging and reaction-time data, proposed that there was a greater conflict involved when people lied about themselves in comparison to when they falsified non-personal information. Duran et al (2020) tested directly if deception success differs depending on the content of the deception: they asked receivers to distinguish truths from lies when senders either spoke about what they did that morning, gave an opinion about a controversial topic, or recounted an emotional event. They found that there was a significant main effect of content type: receivers were worse at distinguishing lies from truths when senders spoke about their morning activities than when they voiced their opinions or recounted their emotional incident. This could be taken as support for the idea that the more personal and emotionally charged the content of the lie is, the easier it is for receivers to catch the lies. However, it is important to note that the senders in this experiment never lied to a receiver or even another person. They simply spoke into a camera once they were assigned to either a lie or truth condition, which in itself raises a different set of issues regarding choice and lack of self-selection, as well as ecological validity, as earlier discussed in Chapter 2.

Our understanding of how deception production and, connectedly, deception detection is affected by type of content is still incomplete. Most deception studies take place in labs and range from lies about feelings to lies about facts, lies that have strong emotional content to those that do not (Hartwig & Bond Jr., 2014); however, cross validation of different paradigms is not a common practice between deception studies using different paradigms. This is an important consideration since, as deception researchers, it is essential to understand how our findings generalise to various content types of lies.

In this study, I use two tasks to elicit deception from senders that differ in the type of content: a task where senders deceive about a picture (Weird Pictures, WP) and a task where they share autobiographical information about themselves (Two truth and a lie, TTL). The aim was to see if deception success differed between the two tasks in a repeated measures design, to gauge the effect that two different paradigms, where each requires senders to deceive about a different subject, will have on their deception production ability. Keeping in line with Frank (2008), where they suggest that lies about pictures that senders see for the first time will not have the same emotional effect as a lie about something more personal, and based on the very limited literature on content of lies (e.g.: Duran et al., 2020), we can predict that:

> **H_2**: Senders will be more successful in deception when lying about impersonal subjects versus more personal subjects, as they are likely to leak more cues when they deceive about personal information. Therefore, their success scores in the WP – where they describe an illustration - will be significantly higher than their scores in the TTL – where they share autobiographical information.

3.1.2. Effects of individual differences on deception production

In Chapter 1, I covered some of the main characteristics of deception. Individual differences in senders play a role in how successful deception is; because Pinocchio's nose does not exist, there is no one-size-fits-all sign that someone is lying. All senders have different cues that they leak, and they leak different cues at different rates. In fact, the variance in how detectable senders are is much higher than the variance in detection abilities of receivers (C. F. Bond & DePaulo, 2008), thus emphasising the influence of individual differences in sender characteristics. These individual differences can stem from many different inherent factors – some cognitive abilities, like mentalizing, and some behavioural traits, such as demeanour and autistic traits.

3.1.2.1. Sender mentalizing ability

In Chapter 1, I reviewed how our understanding of the link between deception and mentalizing stems mostly from child studies. Production of deception is considered a milestone in typical development, marking the presence of theory of mind (ToM), and deception is, in fact, used as a real world test of ToM development (e.g.: Sodian et al., 1991).

Beyond childhood, neuroimaging studies have attempted to map the neural correlates of deception and have found an overlap between deception and mentalizing. A face to face deception study involving functional near-infrared spectroscopy (fNIRS) on both receiver and sender showed that the temporo-parietal junction (TPJ) showed increased activation during deception (versus non-deceptive actions) and the activation in the superior temporal sulcus (STS) was unique to just deception (Zhang et al., 2017). A meta-analysis of fMRI and PET studies showed that the TPJ was one of the regions that showed most activity during interactive deception (Lisofsky et al., 2014). Both the TPJ and the STS have been associated with mentalizing processes (C. D. Frith & Frith, 2006), thus demonstrating that deception production (and detection) likely involves mentalizing.

However, most studies attempting to map the relationship between the two have focussed on the ability to produce deception regardless of actual success, i.e., whether or not mentalizing is related to the *act* of deceiving, rather than the ability of deceiving *successfully*. Only one study to my knowledge has directly investigated the relationship between mentalizing and deception production *success*: Noguchi and Oizumi (2018) applied transcranial direct-current stimulation (tDCS) to the right temporo-parietal junction (rTPJ) and asked senders to provide both a true and false opinion about a highly controversial topic to a receiver, who judged their truthfulness. They found that when judged by an independent group of receivers who watched video recordings of both the true and false opinions, senders' deception success rates fell when their rTPJs were stimulated versus a non-mentalizing related control region.

The deception tasks used in the current study are similar to that used by Noguchi and Oizumi (2018) in two essential ways. One, the senders deceived in an interactive situation, i.e., they were lying to another person who judged their lies in real time (versus lying into a camera or the experimenter); and two, the receivers who judged them watched both veracity conditions and selected which of the two was the lie rather than a dichotomous truth-lie choice of a single stimulus. Therefore, I expect to find a similar relationship between mentalizing and deception production as this previous study did and predict that,

> **H_3**: Individual differences in mentalizing ability will predict senders' deception success, such that higher mentalizing ability will be linked to higher success scores on deception production tasks.

3.1.2.2. Sender demeanour

One of the characteristics pertaining to the sender that has been found to have a profound effect on deception detection – or if considered from a sender's perspective, their ability to produce lies – is sender demeanour. Sender demeanour, as explained in Chapter 1, is how honest or trustworthy an individual appears regardless of their actual veracity (T. R. Levine, 2010). The idea is that a dishonest looking person would be easier to catch when they lie because their veracity will match their demeanour, while an honest looking individual will be harder to catch when the lie because their veracity and demeanour will be mismatched. Up to 98% of variance in deception detection was seen to be explained by sender demeanour across five experiments (T. R. Levine et al., 2011). Levine (2016) tested a classroom full of students (*n* = 27) on both deception production and detection and found that one of the factors that showed the greatest variance was sender demeanour (only behind sender transparency). Semrad et al. (2019), in their meta-analysis, found that sender traits like honestly and believability that contribute to and are used to quantify sender demeanour are fundamental factors influencing deception production. Therefore, there is plenty of support for the idea that how the sender is perceived, how much they are believed, is an important driver of how successful they are in deceiving.

However, as I pointed out in Chapter 2, one of the flaws of traditional deception studies is that every sender usually only provides messages of one veracity: either they tell the truth or they lie. This is problematic, as deception success on these tasks becomes a function of the individual difference in senders' demeanour rather than the senders' actual deception ability. While, in real life, an individual's demeanour probably interacts with their ability to predict success, not controlling for demeanour between senders confounds our measure of their ability. While this is reflective of real life in some scenarios, such as a police interrogation, where the suspect is either telling the truth or lying, in day-to-day interaction, with say, classmates and colleagues, it is likely that a sender sends multiple messages that include both lies and truths. Therefore, the sender's demeanour is likely to influence a receiver's perception of their truths as well as their lies, making a dishonest-looking senders' truths and lies equally unbelievable while making an honest-appearing senders' truths and lies to seem equally credible. For this reason, it is essential to control for demeanour across veracity for every sender and then measure their deception production success, thus giving us a better understanding what factors predict deception success, and if individual differences in demeanour still play a role.

Both the tasks in this study are set up in a way that each sender is judged based on three messages they provide – two true statements and one lie in the case of TTL and one true, one lie, and one double bluff in WP – and the receivers pick out the lie out of the three statements. This allows demeanour to be controlled for across all veracity types for each sender, unlike past studies. However, considering that variance in demeanour has been found to be a stronger factor affecting deception success than actual deception ability of the receivers (T. R. Levine, 2016), we hypothesise that:

> **H$_4$**: Sender demeanour will predict sender deception success: the more honest a sender is rated, the more successful they will be at deception, even though demeanour is kept constant for each sender.

3.1.2.3. Deception production in autism

As explored in Chapter 1, deception production has been studied extensively in autistic children, who have been found to have difficulties in tasks testing either of these abilities (Baron-Cohen, 1992; A. S. Li et al., 2011; J. Russell et al., 1991). Considering that even autistic adults are known to have difficulties in mentalizing (Senju et al., 2009), and mentalizing is thought to be an important factor affecting deception production (Talwar & Lee, 2008), it is a logical leap to assume that autistic adults too will have difficulties in deception. However, there is a dearth of literature when it comes to directly investigating deception production in autistic adults.

Bagnall et al. (2022), in a scoping review, could only identify four studies that looked at deception production in autistic adults. Two of these studies (Davidson & Henderson, 2010; Jaarsma et al., 2012) are qualitative studies exploring the autistic perspective towards deceiving others, and while it is equally important to understand the lived experiences of autistic individuals, these studies do not focus on the psychological mechanism of deception production or the factors that affect deception success in autistic individuals. Of the remaining two studies, Gowen et al. (2008) used a deception production task as a measure of theory of mind measure. While the autistic participants appear to be slightly worse at the penny hiding task (Baron-Cohen, 1992)– a non-verbal game in which senders are required to hide a penny in one hand and keep the receiver from guessing the location of the penny –than non-autistic participants are, no statistics testing this comparison were provided. The other study used a computerised gameplay to test deception production in autistic adults (van Tiel et al., 2021). They found that, in the absence of social norms and requirement to take another's perspective, autistic adults are as likely to deceive as non-autistic individuals are. While the researchers hail this as a strength of the study – isolating deceptive behaviour from social-communicative abilities – it overlooks the crucial point that deception is inherently social and that real world interpersonal deception is very different from gameplay against a computer.

Regardless of the limitations, both these studies attempt to investigate deception production in adults, which is unfortunately otherwise understudied. However, these studies focus on the *frequency* of deception, but just like studies looking at mentalizing and deception production, next to nothing is understood about the *success* of deception and its relation to autistic traits.

As the limited past literature on deception production in autism does not look at success, it also fails to look at how autistic demeanour may affect deception success. Autistic individuals are often perceived to be less trustworthy in general and are thought to lack credibility (Lim et al., 2022). Many of the quintessential autistic behaviours, such as gaze avoidance (Doherty-Sneddon et al., 2013; Hernandez et al., 2009; Volkmar & Mayes, 1990) and repetitive behaviours like rocking and foot tapping (Cunningham & Schreibman, 2008; Jansiewicz et al., 2006) are also perceived by the general population as signs that someone is being deceitful (Global Deception Research Team, 2006). Therefore, it is possible that the presence of autistic behaviour affects deception success in the same was as demeanour: because those who show autistic behaviour could be thought to be less credible, or less honest, they would be easier to catch when deceiving as their demeanour would at these moments match their veracity.

Therefore, along with the idea that the believability of a sender (demeanour) plays such a big role in how effective deception is, it is possible that autistic traits influence how successful a person actually is at deceiving.

This study investigates autistic traits in a sample from the general population, i.e., people who have not been diagnosed with autism. Autism is considered to be one extreme of a neurodevelopmental spectrum, and twin studies have shown that the etiology and rates of heritability of autistic traits are the same in the general population as with those who are clinically diagnosed (Constantino & Todd, 2003; Robinson et al., 2011). Therefore, it is expected that the limited literature surrounding autism and deception production, and our understanding of how autistic behaviours in autistic individuals are perceived, will extend to a sample from the general population reporting a wide range of autistic traits. Considering the novel and untested nature of this paradigm, it is hoped that an autistic traits study will provide a foundation on which future studies with autistic samples can be conducted, helping us advance our understanding on how deception in autism works. Keeping this in mind, I predict:

> **H_5**: Autistic traits will correlate with demeanour, such that senders with higher autistic traits would score lower on demeanour (i.e. more dishonest).
>
> **H_6**: Autistic traits will predict senders' deception success. Autistic traits will affect deception success inversely, i.e. the more autistic traits a sender reports having, the lower they will score on the deception tasks.

3.2. Method

3.2.1. Participants

Forty-two males were recruited through UCL Psychology and Language Sciences online participant database. Thirteen participants only completed the deception tasks and not the other behavioural and questionnaire measures, and therefore were not included in this study. Only 26 of the remaining participants (*Subsample-1*) had at least one usable Two Truths and a Lie (TTL) triplet. Of these 26 participants, a subsample of 17 participants (*Subsample-2*) also had viable Weird Pictures (WP) triplets. All participants were either native or fluent English speakers. The demographic details of all sets of participants are available in table 3.1.

Table 3.1.

Demographic details of all sets of participants

	Full Sample	Subsample-1	Subsample-2
n	42	26	17
Age range	18-26	20-26 years	20-26 years
Age *M*	22.36 years	22.36 years	22.34 years
Age *SD*	1.54 years	1.62 years	1.74 years
English Language (n)	L1: 30 ; L2: 12	L1: 20 ; L2: 6	L1: 13 ; L2: 4

Note. Demographic details for all sets of sample in the study. Subsample-1 is a subset of the full sample who had at least one usable TTL triplet, and Subsample-2 is a subset of Subsample-1 who also had at least one usable WP triplet.

n = sample size, *M* = Mean, *SD* = Standard deviation, L1 = Native English speakers, L2 = Second language speakers of English

3.2.2. Measures

Participants (hereon senders) completed the deception production tasks in pairs. These tasks have been described in detail in Chapter 2, under the title of *Stage 1 – Lie production Study* (section 2.2.1.). Described here are the other measures relevant to this study, i.e. a mentalizing task and an autistic traits questionnaire, which were administered to the senders individually (except for 13 senders who did not complete the individual session due to unavailability or scheduling conflicts). Senders also completed some additional tasks, such as a creativity test, an inhibitory control measure and an IQ test, however these were not analysed as part of this study. All senders also completed a demographics questionnaire.

3.2.2.1. Mentalizing task: Frith-Happé Animations (AT)

To obtain a measure of the participants' mentalizing abilities, the Frith-Happé animations (ToM-AT, Abell et al., 2000) were shown to participants. The task consisted of 12 computer-presented animations in which a pair of triangles, a smaller blue triangle and a larger red triangle, moved around within a boundary. There were four videos of each type:

i. Random: The movements of the triangles were purposeless and arbitrary, and there was no discernible interaction between the two triangles, e.g. floating in space.

ii. Goal-directed (GD): the two triangles interacted so that the actions of one triangle were in response to the behaviour of the other, e.g. one triangle following the other.

iii. Theory of mind (ToM): The interaction between the two triangles was such that it appeared that one triangle was reacting to the other's mental state, e.g. the little triangle mocked the big one by following it, but pretended not to when the big triangle turned to look.

For this study, White et al.'s (2011) administration and objective Multiple Choice Question (MCQ) categorization scoring was used. Participants were given written and verbal instructions: they were to describe what they thought was happening while watching the videos and then answer questions after each video. They were always asked which of three categories (MCQ-categorisation) the video belonged to: “No interaction” (viz. random videos), “Physical interaction” (viz. GD videos), or “Mental interaction” (viz. ToM videos). Participants were also asked about each of the triangles’ feelings (MCQ-feelings) after each ToM videos if the participant had correctly categorized it. Each condition was scored out of four. There MCQ-feelings were scored out of eight – two for each ToM video. Therefore, participants could score between 0 and 20 on the AT.

Participants were shown two practice videos – one GD and one ToM – and were provided feedback. The 12 videos were presented in a pseudorandomized order and all responses were recorded by the examiner on a response sheet. Participant’s verbal descriptions were either audio or video recorded for future analysis (not analysed as part of this thesis).

3.2.2.2. Autistic traits: Autism Spectrum Quotient questionnaire (AQ)

The Autism Spectrum Quotient (AQ, Baron-Cohen et al., 2001) was used to measure autistic traits. The AQ is a 50-item self-report autistic traits questionnaire that records participants’ responses on a 4-point scale (‘slightly agree’, ‘definitely agree’, ‘slightly disagree’, and ‘definitely disagree’) on five different subscales: social skills, communication, attention to detail, attention switching, and imagination.

For each item, participants can score either a 0 or a 1. The scale is traditionally scored such that “slightly agree” and “definitely agree” are scored equally while “slightly disagree” and “definitely disagree” are scored the same. Some items are inversely scored, and after summing scores from all items, a higher overall score indicates higher autistic traits. Participants can score between 0-50, a score of 32 or higher is considered to be a cut-off to distinguish autistic individuals from non-autistic individuals (Baron-Cohen, Wheelwright, Skinner, et al., 2001). This has however been challenged by Broadbent et al. (2013) who found that a cut-off of 29 has lower rates of false positives (<1%) than the original cut-off of 32.

The AQ has been validated in both the autistic and the non-autistic / general populations (Ruzich et al., 2015). It has been found to have high test-retest reliability ($r = 0.95$, Broadbent et al., 2013) and has resulted in a Cronbach's alpha of 0.82 (Austin, 2005) demonstrating high internal consistency as well. In this study, the AQ showed good Cronbach's alpha of .78. The subscales showed lower internal consistency (Social: .61, Attention switching: .67, Attention to detail:.40, Communication: .67, Imagination: .40), this however is not uncommon and similar reliability has been reported in past literature (Hurst et al., 2007).

3.2.3. Procedure

The study was divided in two sessions that either took place on the same day or different days depending on participants availability: a paired session where senders played some deception games with another participant and an individual session where they completed the rest of the tasks and questionnaires. Senders were paid an hourly sum at the departmental rate to thank them for their time and effort. At the start of the study, senders were required to read an information sheet and sign a consent form at the start of the study (and re-consent if the two sessions were on different days). At the end of both sessions, participants were debriefed about the study and its aims. The procedure for the lie production part of the study is explained in detail in Chapter 2, under *Stage 1 – Lie production* (section 2.2.1.).

The videos recorded from the deception tasks underwent strict shortlisting, as listed in Chapter 2 (section 2.2.2.). For the WP, videos were arranged in triplets: a combination of a lie, truth and double bluff. The TTL, were used as is since they were already in triplet form. The shortlisted triplets were then judged by a large number of independent receivers to ensure that the deception production abilities of the senders were objectively recorded and were independent of the subjective deception detection abilities of the other participant they took part with. Receivers also rated the senders on their demeanour. The process of creation of the triplets and demeanour task, along with implementation, is explained in more detail in Chapter 2, under the title of *Stage 2 – baseline study* (section 2.2.2.).

3.3. Results

3.3.1. Bias towards double-bluffing

To explore if senders had a bias towards choosing the less cognitively taxing option to deceive, i.e. double bluff (vs lying), the number of trials on which the sender chose to lie was calculated (*LWP)* for all senders (the full sample), regardless of whether they were successful in deceiving the receivers. To test if senders had a bias, a one-sample Wilcoxon signed-rank test was used as the data was ordinal in nature: *LWP* could range from 0 to 5 (5 trials in WP-P). A median *LWP* of 2.5 would indicate that senders did not have a bias towards either lying or double-bluffing. A median significantly higher than a test value of 2.5 would denote a bias towards lying, and conversely, a median lower than 2.5 would indicate a bias towards double-bluffing.

The sample median of 2 (*IQR* = 1) was not found to be significantly different from the test value of 2.5, $Z = 147$, $p = .111$. This means that senders did not have a significant bias towards double bluffing more often than lying in WP.

3.3.2. Deception ability in different tasks

To calculate the deception ability of the senders, every sender in the Subsample-1 sample was awarded 0 if a receiver correctly caught their lie and 1 if a receiver answered incorrectly for each triplet. Responses were averaged across all receivers in all batches of the baseline study for each sender; a weighted average was calculated as different batches in the baseline study had different numbers of receivers. Therefore, for each sender, a score for successful deception for both the WP and the TTL – S_{WP} and S_{TTL} respectively – was calculated. Scores ranged from 0 to 1, a lower score indicating lower deception ability.

To test if deception ability differs depending on what one deceives about, task performance on TTL was compared to WP for Subsample-2. A Shapiro-Wilk test of normality showed that all data were normally distributed. A paired sample t-test to examine if senders differed in their ability to lie successfully in the WP and TTL showed that senders in Subsample-2 were significantly more successful at deception in TTL ($M = 0.67$, $SD = 0.11$) than in WP ($M = 0.57$, $SD = 0.15$); $t(16) = 2.63$, $p = .018$, $d = .637$.

3.3.3. Associations between deception success, autistic traits, mentalizing, and demeanour

Demeanour score was calculated by taking a weighted average of all the demeanour ratings from all receivers for each sender. To assess whether deception success correlates with mentalizing ability, demeanour, and autistic traits, Pearson's correlations were calculated for each task between the senders' success and their total score on the AT, Demeanour scores, and the AQ. Deception success scores in neither of the tasks showed any significant correlation with either the AT, their demeanour scores, or the AQ (Table 3.2). Further, a Pearson's correlation also showed that there is no evidence to suggest that there is an association between autistic traits and demeanour (Table 3.2).

Table 3.2.

Correlations between deception success, mentalizing, and autistic traits, and demeanour

	M	*SD*	Possible range	Actual range	1	2	3	4
1. AQ	18.92	6.40	0-50	6-30				
2. AT	13.50	2.53	0-18	9-17	-.108			
3. S_{WP}	0.57	0.15	0-1	.29-.86	.038	-.455		
4. S_{TTL}	0.65	0.13	0-1	.35-.88	.076	-.118	.310	
5. Demeanour	4.16	0.69	1-7	2.92-5.51	.156	0.66	-0.034	0.152

Note. M = mean, *SD* = standard deviation, AQ = Autism Spectrum Quotient, AT = Frith-Happé Animations, S_{WP} = deception success score in Weird Pictures, S_{TTL} = deception success score in Two Truths and a Lie

All correlations pertaining to S_{WP} were carried out on Senders-3 (n = 17), while all other correlations were tested on Senders-2 (n = 26).

3.3.4. Predicting deception production success

Considering that, none of the predictor variables, i.e. autistic traits (AQ), mentalizing (AT), and Demeanour, showed any correlations with the outcome variables, i.e. deception success in the novel deception tasks S_{WP} and S_{TTL}, the planned hierarchical regression analysis was not performed as it was no longer expected that the predictor variables would explain variance in the outcome variable.

3.4. Discussion

In this chapter, I investigated non-autistic senders' deception production ability. Senders took part in two novel deception production tasks – the Weird Pictures (WP) and Two Truths and a Lie (TTL) – along with other measures assessing their mentalizing ability and autistic traits. An independent batch of receivers attempted to detect the senders' lies in a separate study, thus providing objective measures of sender deception production success for both WP and TTL. The receivers also rated the senders on how honest or dishonest they appeared to be, regardless of their veracity, providing a measure of senders' demeanour.

Contrary to my hypothesis, senders did not show a bias towards double bluffing over falsifying/lying outright (H_1). While there was a difference in senders' success in TTL and WP, the difference was in the opposite direction to what was predicted (H_2). Unfortunately, unlike what I expected, none of the associations between the success at the deception tasks and the individual difference variables – mentalizing (H_3), autistic traits (H_4), demeanour (H_6) – were significant, and demeanour and autistic traits were not found to be related either (H_5).

3.4.1. Cognitive effort and choice of deception

This study used two novel deception production tasks – the WP and TTL – both created to combat some of the shortcomings of past deception paradigm, such as the senders' lack of choice about how and when to deceive (detailed in Chapter 2). Senders in WP were given a choice between lying outright, i.e., falsification, and double bluffing to deceive the receiver.

It was predicted that senders would choose to double bluff over lying, as theoretically, lying likely uses more cognitive process – inhibitory control and increased working memory load – in comparison to double bluff, and that senders would have a bias towards the option that is thought to be the less cognitively taxing way of deceiving. However, the senders in our sample did not show such a preference to double bluffing.

This aligns with some of the neuroimaging literature that suggest that both lying and double bluffing use common areas in the brain and show similar levels of activation in those areas (Kireev et al., 2017; Volz et al., 2015; Zheltyakova et al., 2020). Contrarily, functional near-infrared infrared spectroscopy (fNIRS) analysis by Ding et al. (2014) demonstrated higher activation of the right middle frontal gyrus during lying versus manipulative truth telling. However, a higher activation does not definitively mean a higher cognitive load during lying, which would in turn translate into a behavioural preference towards double bluffing. Therefore, it cannot be said for certain if the findings of this study refute those by Ding et al. (2014).

One explanation to the lack of bias could be a possibility that double bluffs do actually involve as much inhibitory control as lying does. Carrión et al. (2010) showed in an electroencephalography, that the event related potentials associated with cognitive control were identical for both double bluff and lies, and that they were significantly different from the ERPs of truth-telling. Other studies have also since then found similar results and further suggest that increased demand for cognitive control is associated with deceptive intent, regardless of whether the sender deceives by lying or by being honest (Sai et al., 2018). Further, it is suggested that this this increased need for cognitive control does not stem from supressing the urge to tell the truth but because of the conflict experienced by the senders on considering the mental state of their receivers (Leng et al., 2019).

An alternative explanation for the similar preference towards lying and double bluffing could be that it is an artefact of the task design. The fact that the participants showed an equal preference towards lying and double bluffing can be taken as an indication that having the autonomy to choose *how* they deceived resulted in them choosing to lie as often as double bluff and being more invested in their lies than if they lied on instruction. It supports the idea that choice and monetary rewards, which were both measures introduced in these novel tasks to incentivise deception, may possibly increase motivation to lie. This would mean that the absence of bias is not because lying is *not* considerably more mentally taxing than double bluffing, but because participants are more motivated to win the 'game' and therefore strategize differently than they would if there was no motivation and only cognitive effort to base their decisions on. It is possible that participants used strategy for the overall task rather than for each trial separately – they lied and double bluffed equally because that made them more unpredictable and gave them a strategic edge in the overall game against the receiver they were partnered with. It will be interesting to see if any possible additional cognitive effort expended on lying manifests as lower success rates for lying than double bluffs, and future studies could focus on how types of lies and a potential difference in their respective cognitive load affects success.

3.4.2. Cognitive effort and content of deception

It was expected that deceiving about pictures would require lesser cognitive effort due to the impersonal nature of the pictures that will in turn mean that senders will not have any emotional attachment and investment in their deception, versus sharing and deceiving about autobiographical information (Frank, 2008; Nuñez et al., 2005). Therefore, it was predicted that senders would be better at the WP than TTL, as they would leak fewer cues when deceiving in WP, making their deception more successful. Contrary to this, senders were found to be significantly better at deceiving in TTL than WP, with a moderate to large effect size.

The reasoning behind the expectation that senders would leak more cues in TTL stems from the idea that lies with higher stakes puts a bigger cognitive load on the sender (Frank & Ekman, 1997) and that people undergo more conflict when falsifying information about themselves versus impersonal information (Nuñez et al., 2005). Duran et al (2020) found that senders were able to successfully deceive when they described non-emotional topics like their morning routine, and were only significantly worse when they had to deceive about an emotional experience or when they had to deceive about something that may give rise to internal conflict – lying about a controversial topic. However, it is possible that the lies in TTL – about senders' summer plans, hobbies, their weekend routine, and their future goals – while of a more personal nature than those in WP, were not enough to elicit a sufficiently high emotional response to cause them to leak more cues. It is also possible that gamifying the deception tasks – such that the senders were rewarded for successfully deceiving the receiver in front of them reduced the conflict that would have otherwise arisen when falsifying autobiographical information.

However, this still does not explain why TTL performance was significantly better than WP. DePaulo et al (2003) report in their meta-analysis, that senders who construct a lie from scratch show more signs of cognitive effort. They suggest that the effort of fabricating with a complex lie leaks out as various physical cues such as reduced expressiveness, blinking, speech disturbances, etc. The WP required senders to view a picture and then, if they decide to lie (instead of double bluff), potentially imagine an alternate image that they could describe. While some senders chose to borrow ideas from the picture they had viewed, these lies were discarded during the shortlisting process, as they were deemed similar to the actual image and that was considered a drawback when creating the deception detection paradigm (Stage 2 – Baseline study, Chapter 2). On the other hand, the TTL did not go through a shortlisting based on how similar the lie was to the truth (as the ground truth was not known) and one of the common responses to the debrief interview after the TTL was that their lies were borrowed from real life experiences. Senders reported that they changed some aspect of the truth and supplied it as a lie; this could be changing something insignificant (e.g.: talking about how they love Cricket, when they actually do not) or borrowing from someone else’s experiences (e.g.: passing off their friend’s holiday plans as their own). Therefore, the fact that senders had to be more creative and fabricate their lies from scratch in WP, unlike the TTL, could have led to more cue leakage in the WP, making them worse at deceiving in that task, contrary to what was predicted.

Therefore, the results could be a by-product of the exclusion criteria to shortlist videos for the WP. Alternatively, there is a chance that it could simply be that the ground truth was not known in TTL, and that senders were not honest about which of the statements was actually the lie. This, however, is unlikely, as explained in Chapter 2.

3.4.3. Autistic traits and demeanour

As autistic individuals are perceived to be less credible (Lim et al., 2022), it was hypothesised that senders who reported having more autistic traits would be judged to be more dishonest. However, this was not found to be true, as there was no correlation between their AQ scores and the mean demeanour scores given by the independent receivers.

The simplest explanation to this is that the present sample of non-autistic individuals did not show enough autistic behaviour to be subjected to the same negative impressions as diagnosed autistic individuals do. In fact, none of our senders reached the AQ cut-off (32) that is thought to distinguish autistic individuals from the general population (Baron-Cohen, Wheelwright, Skinner, et al., 2001). As Sasson and Bottema-Beutel (2022) point out, studies of autistic traits cannot be equated with studies in an autistic sample, and the opposite can also be taken to be true. Simply because autistic individuals are perceived more negatively or as less credible, does not mean that non-autistic individuals (senders) reporting high autistic traits will be subjected to the same attitudes from other non-autistic individuals (receivers). This is even more relevant because many of the ‘autistic’ traits measured by self-report questionnaires like the AQ, co-present in autism, whereas they may occur in isolation in the general population, which does not necessarily reflect autistic symptomatology (Mottron & Bzdok, 2020). Therefore, it is understandable that results that would be expected in an autistic sample do not extend to a non-autistic sample.

The other factor that could explain the results is the use of the AQ and its suitability for this study, considering our sample was non-autistic. The AQ, which was developed primarily to distinguish autistic adults with normal intelligence, may not be as sensitive in assessing the degree of autistic traits in neurotypical adults of normal intelligence as previously believed (Lundqvist & Lindner, 2017). The AQ has also found to have weaker criterion validity

(Ingersoll et al., 2011), and internal consistency (Nishiyama et al., 2014) when testing for autistic traits in non-clinical populations in comparison to other measures of autistic traits such as the Social Responsiveness Scale for Adults (SRS-A; Constantino et al., 2003). Furthermore, the AQ, being a self-report measure, does not correlate with observer-rated measures of autism such as Autism Diagnostic Observation Schedule (ADOS; Roestorf et al., 2018), and autistic traits may be underestimated by self-report (Möricke et al., 2016). Another issue with the AQ is that participants were aware that the study was autism-related, which could have influenced how some of them answered the questions in the AQ.

3.4.4. Predicting deception production success

Considering the body of literature evidencing a relationship between mentalizing and deception production, especially in autistic children, and the emerging importance of demeanour in deception detection, and therefore production, I proposed that senders' mentalizing ability, autistic traits, and demeanour would predict the variance in their success in the deception tasks. Unlike what was hypothesised, none of the predictor variables were found to be associated with deception success.

3.4.4.1. Mentalizing and deception success

It was found that in this sample of neurotypical male adults, success on either of the deception tasks was not associated with performance on the mentalizing task. While our finding contradicts previous studies in children, which show deceptive behaviour to be associated with false belief task (FBT) performance (e.g.: Talwar et al., 2007), the results found here are not unprecedented. Wang, Zhu, & Wang (2017) found that children's strategic deception of the examiner about a transgression during a temptation-resistance paradigm did not correlate with change of location and unexpected contents FBT. Newton, Reddy & Bull

(2000) found that the amount and variety of day-to-day deception by children was not linked to their FBT performances. However, it should be noted that these studies record instances of deception and not successful deception.

That said, the findings of this study do not align with those of the previous study that looks at success of deception in adults – Noguchi and Oizumi (2018) recorded tasks with an interactive element, and played both lie and truth to an independent set of receivers, just like the tasks in this study. They found a causal effect of mentalizing on deception production success. However, it is important to note that of the three deception production tasks the senders completed, success on only one task was affected by stimulation of mentalizing associated brain regions. The opinion task, where senders provided true and false opinions on a controversial topic, was the only task that involved any degree of internal moral conflict, probably arising from going against a personal belief for monetary reward. This could in turn explain why stimulation to the TPJ, which has been associated with mentalizing based moral conflicts (L. Young & Saxe, 2009), influenced their deceptive behaviour in the opinion task, and not on the other two deception tasks. Both these other tasks in Noguchi and Oizumi (2018)'s study – a shock test where the sender deceived about a mild electric shock or a beverage taste test – did not have a component that would induce a moral conflict in the senders, and neither did the two tasks used in this current study. This could possibly explain why mentalizing is not correlated to the senders' deception success in the current study.

Another justification for the absence of correlation between mentalizing and deception production success is that the Frith-Happé animated triangles used in this study are traditionally used to detect mentalizing difficulties in an autistic population, and to establish group differences between clinical and control groups. While there have been studies that use it to draw correlations and comparisons between ToM and other variables, these studies have much larger sample sizes than the present study (e.g.: Jones et al., 2011; Koelkebeck et al., 2017). Moreover, White et al (2009) found that for the objective scoring they devised (and which was used here), autistic participants performed worse than the non-autistic participants only on the identification of ToM animations and the MCQ-Feelings. It is therefore possible that this task is not sensitive enough to pick up individual variability in a small sample of non-autistic participants. One possible solution to this is to examine the verbal descriptions of each video provided by the participants (which were recorded but not analysed as part of this study) as they were relatively unprompted and may be more sensitive to ToM variability within the sample.

3.4.4.2. Demeanour and deception success

Contradicting my prediction, demeanour did not correlate with deception success in the present study. While this contradicts predictions made for this study and past literature (T. R. Levine, 2016; Semrad et al., 2019), there are two alternative explanations that could possibly explain these findings, one is a potential limitation of the demeanour task and the other is a definite strength of the study design.

The first potential reason why no relationship between demeanour and deception success was observed could be the way demeanour scores were obtained. The construct of demeanour as per Levine (2011) goes beyond a one dimensional measure of the senders behaviour. More specifically, they outline and validate in multiple samples 11 different criteria and derive an index of honest demeanour from the average of these 11 unidimensional criteria. In comparison, the single scale asking the independent receivers "*how honest does this person appear?*" is likely not sensitive enough to capture the complex nature of sender demeanour. Additionally, the demeanour video stimuli in this study were devoid of any sound – a necessity arising due to the fact that the senders in the videos described their deception strategies during the debrief session from which the demeanour videos were extracted and demeanour ratings, by definition, need to be unrelated to deception or veracity. Considering that two of the 11 Levine criteria are related to verbalisations – vocal uncertainty and verbal uncertainty – this further lessens the validity of the one-scale demeanour measure of this study. That said, it cannot be disregarded that Semrad et al (2019) have demonstrated that results from basic personality questionnaires quantifying just believability and credibility may be successful in predict deception ability.

The other reason for this lack of association could be due to the experimental manipulation introduced in this study – controlling for demeanour across veracity. An honest appearing liar is more successful because when the receiver is simply making a lie / truth judgement, they base their judgement on the sender's demeanour rather than the actual veracity of message. Contradictorily, when receivers are required to pick the lie from multiple messages from the same honest appearing liar, the degree to which an individual appears honest no longer affects the credibility of the lie alone because all messages appear equally credible. Therefore, no effect of individual differences in demeanour on deception production success is observed.

Demeanour was kept constant across veracity in this study because, as detailed in Chapter 2, this study – the lie production and baseline studies – was in part a means to develop two new deception detection tasks, ones that were purer measures of receivers' deception *detection* without the confounding effects of sender-based characteristics, such as sender demeanour. If the lack of association between demeanour and deception production indicates that the effects of demeanour are nullified by holding it constant across veracity types, this is a definite strength of the deception detection task that I sought to create (Chapter 2).

While effects of demeanour have been thought to have influenced deception success for decades (T. R. Levine, 2010; Riggio et al., 1987; Van Swol et al., 2015), it has emerged as a factor that overshadows all other factors that affect deception success (T. R. Levine et al., 2011). However, if keeping demeanour constant across veracity limits its effects, this has consequences on how interpersonal deception is studied. When people are lied to by someone that they come across multiple times (versus in a criminal setting where the accused's veracity likely needs to be judged in the context of a singular instance), receivers do not have the advantage of judging the sender based on their general demeanour. Therefore, there is a need to move beyond just looking at demeanour as the most significant predictor of deception success.

3.4.4.3. Autistic traits and deception success

The findings in the present study did not demonstrate a relationship between autistic traits and deception success. While there are hardly any studies on deception production in autistic adults, this goes against the child literature in this field that suggests that autistic individuals have difficulties in producing deception (Baron-Cohen, 1992; A. S. Li et al., 2011; J. Russell et al., 1991). However, it is possible that the problems that autistic children have in deception production stem from difficulties that adults do not have. Alternatively, it is possible that just as autistic adults have been found to deceive as *frequently* as non-autistic adults (van Tiel et al., 2021), adults who report higher autistic traits are similarly as *successful* at deceiving receivers as those adults who report fewer autistic traits, thus explaining why autistic traits and deception success show no associations.

That said, it is important to reiterate that the senders in this study were not actually autistic. Therefore, it is possible that the autistic traits in the study were not substantial enough to begin with to affect deception success. Little is known about deception-production in adults showing autistic traits; however, Kunihira et al. (2006) found that autistic traits in the general population did not correlate with executive functioning (see also: Maes et al., 2013; cf. Christ et al., 2010), unlike that in autistic samples (R. M. Joseph & Tager–Flusberg, 2004; Van Eylen et al., 2015). Considering deception-production requires multiple cognitive functions (Gombos, 2006), it is therefore understandable that autistic traits in this non-autistic sample did not correlate with deception production ability.

It is also noteworthy that one of the premises for predicting an association between autistic traits and deception-production was that autistic individuals have mentalizing difficulties that underlie their social difficulties (Senju et al., 2009), and that mentalizing ability is in turn associated with deception. Since no significant relationship was found between mentalizing and deception in this sample, it logically follows that the expected negative correlation between autistic traits and deception was not observed either.

Alternatively, it is also possible that the task design was responsible for the lack of correlation between autistic traits and deception success similar to how it was speculated that controlling demeanour across veracities for each sender negated the effects of demeanour on deception success. Any autistic behaviours such as gaze avoidance and foot tapping that are common to both autism (Cunningham & Schreibman, 2008; Doherty-Sneddon et al., 2013) and deception (Global Deception Research Team, 2006), would have been exhibited in all veracity types for each sender, thus not making the sender any less or more credible in any one veracity. Therefore, the disadvantage that a sender could have likely had due to having high autistic traits is neutralised because the behaviour is equally present in all videos the receiver would have watched of them, thus not making the sender any worse at deception than they would be without their autistic behaviours.

3.4.5. Limitations of the current study

The current study is not without its shortcomings. Task related limitations – such as lack of confidence in ground truth for the TTL, sanctioned nature of lies, and possible fatigue in the independent receiver – and their trade-offs have been discussed in detail in Chapter 2.

One limitation of this study that has been highlighted throughout is the fact that the current sample was made up of just non-autistic adults and that limits the generalisability of the findings to the autistic population. This was not entirely by design – the initial plan for this study included autistic as well as non-autistic senders, however due to the pandemic and lockdowns, it was not possible to collect data from autistic individuals.

It also must be considered that the deception production study was primarily a step in the production of the deception detection stimuli. This also severely limited the sample size for the study, and it is possible that the sample was not large enough to pick up any possible associations between our variables.

It is, however, possible that that relationships between the variables being investigated in this study would be better demonstrated in a larger sample including autistic participants, as having "autistic traits" in isolation are not the same as the experience of and the difficulties associated with actually being autistic (Sasson & Bottema-Beutel, 2022). Therefore, given the paucity of research regarding deception production in autistic individuals, and deception success specifically, it is important that future studies investigate this further. This is especially vital as autistic individuals describe using deception, specially verbal deception, as a means of self-protection (Jaarsma et al., 2012), especially in social situations where there is a possibility of facing stigma (Davidson & Henderson, 2010).

3.5. Conclusions

In this study, I set out to investigate how cognition, individual traits and characteristics affect deception production.

It was predicted that people in general have a bias towards choosing the least cognitively effortful means of deceiving their partners, and therefore would choose to double bluff, i.e., tell the truth while pretending to lie, more often than outright lying or falsification. However, no such bias was observed in this study, which could be because, unlike what was assumed on the outset, double bluffing may carry as much cognitive load as falsification does. Alternatively, it could be due to the fact that the increased motivation to deceive well, introduced by choice and extrinsic monetary reward, encouraged participants to falsify as frequently as double bluffing. The lack of bias could also potentially be strategic – senders lied and double bluffed intermittently to be unpredictable in order to gain an overall edge in the ‘deception games’.

The second aspect investigated in this study was the effect of type of lie and its associated cognitive load on deception success. It was predicted that lies with more personal information would be deemed as having higher stakes and would increase the amount of cues the sender leaked, making them worse at deceiving. It was surprising then that deception success on the task involving the autobiographical information was found to be higher than the task involving impersonal picture descriptions. This was chalked up to the possibility that deception about pictures involved creating lies from scratch whereas lying about autobiographical information involved drawing inspiration from real life, which made the former more effortful for the senders.

Finally, the effects of sender’s mentalizing abilities, demeanour, and autistic traits on deception production were also investigated. Contrary to expectations, deception did not show any discernible association with any of the three. This could be because of the small sample size, or the unsuitability of the mentalizing and/or autistic traits measures used.

The relationship between autistic traits and deception success, in the context of mentalizing and demeanour, needs to be investigated further in the future, preferably in a larger sample of autistic participants, who have clinical diagnoses, unlike the non-autistic sample in the present study. Using more sensitive tests for mentalizing, such as the subjective descriptions in the Frith-Happé animations tasks, and for autistic traits, and more comprehensive scales for demeanour, might also help illustrate the relationship between autistic traits, mentalizing, demeanour and deception.

Chapter 4. Development of Deception Detection

4.1. Introduction

Television shows like *Lie to me* (Baum et al., 2009–2011) would have people believe that certain skills, like reading micro-expressions, make people expert lie detectors. This may be true with the help of visual search computer programs and algorithms (Su & Levine, 2016) and some people may be good at detecting micro-expressions by a combination of natural talent and training, but this is not something that most people are capable of (Frank & Svetieva, 2015). On average people are only accurate 54% of the time when detecting lies, which is still significantly above chance (C. F. Bond & DePaulo, 2006). In the last few decades it has been suggested that deception detection success is mostly based on the sender characteristics (e.g. C. F. Bond & DePaulo, 2008), however the existence of good lie detectors and bad lie detectors (e.g. Mann et al., 2004) show that there must be individual differences in receiver ability that contribute to deception detection as well. So, what other individual abilities, apart from the rare ability to read micro-expressions, help people detect lies, especially in day-to-day scenarios? Is deception detection something certain people are good at while others are not? Can this ability change over time?

Earlier in Chapter 1, I offered that while there are no set cues that are common to all liars (Hartwig & Bond Jr., 2014) and deception detection may not be dependent on factors like education and experience (Aamodt & Custer, 2006), there could be cognitive factors – namely mentalizing – that drive individual differences in deception production and detection. In fact, deception and mentalizing are so entwined, that the former is considered a marker of the latter (Hala et al., 1991). Based on this line of thought, it can be proposed that any changes in mentalizing will predict changes in deception detection.

In this chapter, I first verify Klindt et al.'s (2017) claims that mentalizing develops with age. Considering autistic individuals are thought to have difficulties in mentalizing (Senju et al., 2009), but autistic adults also have been found to pass traditional false-belief mentalizing tasks (Bowler, 1992; Frith, 2004), I will reconfirm if mentalizing is affected in autistic individuals and additionally examine if they show the same developmental trajectory as non-autistic individuals do.

I then go on to investigate if deception detection can be predicted by mentalizing, if it develops with age in the same way as mentalizing is thought to, and if deception detection is affected in autistic individuals like mentalizing is thought to be. Finally, I look at what impact weakened deception detection can have on the quality of people's lives – especially on how vulnerable to peer-victimization it may make them, and if this in turn affects their mental health.

4.1.1. Development of mentalizing and autism

In Chapter 1, I established how mentalizing is thought to change and develop with age. Blakemore (2012) maintains that the brain continues to develop structurally and functionally beyond childhood. Mentalizing related brain regions continue to show variable activation with development even after early childhood, for example, adolescent participants in comparison to adult participants have been found to show activation of a larger volume of the Medial Prefrontal Cortex (mPFC) as well as heightened activation within parts of the mPFC (Burnett et al., 2009). While this indicates a definite change in the way the brain processes social information with age, what this means in terms of actual ability and behaviour, though, is still under investigation.

Especially robust evidence of not just change, but *directional development* in mentalizing ability comes from a multi-paradigm crowdsourced online study (Klindt et al., 2017). Using multiple lower order mentalizing tasks, such as traditional first order false belief tasks, and higher order mentalizing tasks, such as an adaptation of the Frith-Happé animations task (S. J. White et al., 2011), Klindt et al. (2017) showed that lower order and higher order ToM have different developmental trajectories – lower order ToM develops faster than higher order ToM. More importantly, they demonstrated that mentalizing ability not only continues to develop beyond early childhood, but that it peaks in early adulthood (around mid-20s).

Klind et al.'s (2017) study has many strengths, using a range of tasks being one of them. While they boast of an impressive sample size – over 30000 participants between the ages of 10 to 85 years – the majority of participants took part in only one of the tasks in the battery, and therefore their samples in the youngest age groups in every task were much smaller than the average sample size – under 100 participants in each task. Add to that the small (albeit significant) effect of age on mentalizing they found, it is prudent that the results be replicated before being considered as conclusive proof of development of mentalizing – especially in late childhood / adolescence up till early adulthood.

There is plenty of literature suggesting that autistic individuals have difficulties in mentalizing (e.g.: Senju & Csibra, 2008; reviewed in detail in Chapter 1), but research on the development of mentalizing in autism beyond childhood is scarce. Bal et al. (2013) sought to test if mentalizing developed differently in autistic and non-autistic individuals (7- to 17-year-olds) utilizing one of the tasks Klindt et al. (2017) used – the Frith-Happé animations. They found that while mentalizing ability developed with age in non-autistic participants, this relationship was weaker for the autistic participants. Moessnang et al. (2020) used the same task and discovered that while age was a significant predictor of mentalizing ability, this did not differentiate autistic from non-autistic individuals. The mixed results give rise to the need for further investigation into the development of mentalizing ability in autistic and non-autistic individuals.

The present study aims to use an online adaptation of the Frith-Happé animations task (Livingston et al., 2021) to re-examine if there is a difference in how mentalizing ability develops in autistic and non-autistic individuals in a cross-sectional sample that is more evenly distributed throughout the developmental range (11-30 years). Taking cues from past studies showing a clear difference in socio-cognitive abilities between autistic and non-autistic individuals (see Baron-Cohen, 2000a for review) and the development of mentalizing ability with age (Bal et al., 2013; Klindt et al., 2017; Moessnang et al., 2020), it was predicted that:

> H_1: There will be a significant difference between autistic and non-autistic participants in their ToM abilities, i.e., their performance on the Frith-Happé animations task.
>
> H_2: Mentalizing abilities for both autistic and non-autistic individuals will positively correlate with age, i.e., the older the participants, the better their performance on the mentalizing measure.
>
> H_3: There will be a significant interaction between diagnostic group and age, i.e., while both groups will show positive correlation with age, the non-autistic group's mentalizing ability will show a stronger association with age than that of the autistic group. In other words, mentalizing ability will be observed to develop more in the non-autistic sample with age than in the autistic sample.

4.1.2. Mentalizing and deception detection

Earlier, I mentioned how deception and mentalizing are thought to be so intwined that attempts at producing deception are often taken as signs of a developing ToM (Sodian, 1991), and that ToM training leads to children lying more often (Ding et al., 2014). Our understanding of the relationship between deception and mentalizing comes primarily from deception *production*. While intuitively, it can be extrapolated that deception *detection* too would require mentalizing – an understanding that someone else intends to plant a false belief in us seems crucial to deception detection – the relationship between deception detection and mentalizing has hardly ever been empirically investigated.

Sowden et al. (2015) showed that there is a link between processes that contribute to mentalizing and deception detection ability: transcranial direct current stimulation (tDCS) to the right temporoparietal junction (rTPJ) – an area that has been previously linked to mentalizing (C. D. Frith & Frith, 2006) – improves deception detection in situations where the senders' opinions are contradictory to the receivers. The authors attribute this improvement in context-specific deception detection ability to an improvement in the receivers' ability to separate self and others' representations, which is a key process in belief attribution and mentalizing (Spengler et al., 2009). This nevertheless does not evidence a direct link between deception detection and mentalizing abilities.

A relationship between mentalizing ability and recognition of suspicious or 'dodgy' behaviour has also been established – Brewer et al (2018) had participants listen to vignettes of scenarios that may or may not end in a crime being committed and asked them to respond if and when they realised that something suspicious was happening. Participants who scored higher on the mentalizing measure, the Frith-Happé animations (S. J. White et al., 2011), were quicker to correctly judge suspicious activity, albeit with small effect sizes. Brewer at al. (2018) cite their inspiration to investigating this unique relationship to the high incidence of autistic individuals who are naively involved in criminal scenarios (Brewer & Young, 2015) but, even though they agree that ToM abilities may be more variable in autistic individuals, they do not test autistic individuals in their study.

A more direct relationship between mentalizing and deception detection has only been tested twice, in my understanding. Stewart et al. (2019) found that truth detection and deception detection use different cognitive processes, and that deception detection ability is linked to emotional intelligence and cognitive theory of mind. Williams et al. (2018) used the Cheating Tapes (T. R. Levine, 2007–2011) – a well-established deception detection paradigm (e.g.: Barnicle, 2021; Curtis et al., 2021; Zloteanu et al., 2021) – and correlated performance with two mentalizing measures: the Frith-Happé animations (Abell et al., 2000) and the Reading the Mind in Eyes Task (REMT, Baron-Cohen, Wheelwright, Hill, et al., 2001). The RMET, however, has been criticized for not being a tool for measuring pure ToM, but also emotion recognition (Oakley et al., 2016). Williams et al (2018) do not report finding a significant relationship between mentalizing and deception detection. It is noteworthy though that they do observe a weak correlation that is marginally significant (p = .09) on at least one of the conditions (not specifically mentioned in their manuscript) in the deception task. This leaves open the possibility that a bigger sample may enable us to uncover a potential relationship.

Neither Stewart et al. (2019) nor Williams et al. (2018) investigate the relationship in autistic individuals, even though the latter tested deception detection in autistic individuals (study 2). In fact, no study has attempted to study the link between deception detection and mentalizing in autistic individuals, despite the plethora of literature pointing towards autistic children's inability to produce deception due to ToM difficulties. Considering that a reduced ability to detect lies in autistic individuals is based on the premise that they are thought to have difficulty mentalizing, this current study seeks to correct this oversight. As part of the deception detection battery this study uses the same lie detection task used by Williams et al. (2018) – the Cheating task (T. R. Levine, 2007–2011) – and uses one of the mentalizing measures they used as well – the Frith-Happé animations, albeit one adapted for online use

(Livingston et al., 2021) to test an autistic sample and a larger non-autistic sample. It is therefore predicted that,

> H_4: Deception detection ability will show a positive relationship with mentalizing ability, for both autistic and non-autistic individuals.

4.1.3. Development of deception detection

In Chapter 1, I explained how studies focussing on the development of deception beyond childhood are extremely scarce, and the few that do investigate it have looked at deception production instead of detection.

There have been a handful of studies that have attempted to map the effects of development from childhood to adulthood in deception detection. An old study investigated 8- to 33-year-olds and found that older participants were seen to be better at decoding non-verbal cues (DePaulo & Rosenthal, 1979); this could possibly translate into better ability to detect deceptive cues as well. DePaulo et al. (1982) attempted to tests deception detection more directly and found that older adolescents and young adults are able to distinguish between genuine and feigned affect better than younger adolescents and pre-adolescent receivers. Senders were recorded describing someone they genuinely liked or disliked (honest condition) or describing someone they liked whilst pretending they disliked them (deceptive condition) or describing someone they disliked as if they liked them (also deceptive condition). Senders in 6 age groups – sixth graders (M_{age} = 11.5 years), eighth graders (M_{age} = 13.6 years), tenth graders (M_{age} = 15 years), twelfth graders (M_{age} = 17 years), and college students (M_{age} = 18.1 years) – then judged the affect of these senders. It was found that while the younger children rated the deceptive senders' affects as whatever their feigned expression was, only the oldest three groups successfully picked up the deception by rating the deceptive senders' affect as 'mixed'. However, the results must be considered in light of the fact that the younger receivers were also worse at assigning the correct affect (like or dislike) to the senders even when their expressions were genuine. Therefore, it is possible that the effect of age on deception detection observed here is an artefact of a better understanding of affect generally in the older participants.

More recently, Shaw and Lyons (2017) found age to be a significant predictor of deception detection when 16-67 year-olds were asked to judge public appeals for help from relatives of murdered or missing individuals. Half of these appeals were false appeals (i.e., the person appealing was later found to be involved in the murder or disappearance of their relative). Overall, older participants emerged as better detectors in this study. Half the participants were given a list of previously identified cues to deception before the task, and it is critical to note that the group with cues was significantly older than the group without cues, which could have potentially contributed to the enhanced deception detection abilities observed for the older participants. Therefore, it is not certain that deception detection actually improved with age.

Considering that mentalizing is thought to be one of the cognitive abilities contributing to deception, and mentalizing is thought to develop beyond early childhood (Klindt et al., 2017), this study aims to investigate the development of deception detection in a sample that captures the development of mentalizing: from late childhood (pre-adolescence) through adolescence till early adulthood. Based on the findings of the few studies that have touched upon development of deception detection in the past, and on the studies mapping the developmental trajectory of mentalizing, I hypothesize that,

> H_5: Age will predict deception detection ability i.e., the older the participants are, the better they would be at detecting deception.

4.1.4. Deception detection and autism

As development of deception skills is taken as a mark of developing mentalizing (Hala et al., 1991), deception in autism, especially in the childhood, has been investigated before. However, these have been mostly limited to deception *production* (E.g.: Baron-Cohen, 1992; A. S. Li et al., 2011; J. Russell et al., 1991; Sodian & Frith, 1992). Studies that do touch upon deception detection in autistic individuals, children and adult alike, have only done so to in context of autistic individuals' understanding of complex social constructs like such as sarcasm and jokes (Leekam & Prior, 1994; S. White et al., 2009).

In one of the very few studies that have looked at deception detection in autistic adults, Sowden et al. (poster abstract, 2018) demonstrated that autistic traits were linked to how capable receivers were in detecting lies, especially in situations where the senders' opinions (either true or false) were contradictory to the receivers' own views. They found that higher the autistic traits, more difficult it was for participants to detect lies where there was a self-other contradiction. Although they tested autistic individuals, they do not compare autistic and non-autistic individuals directly; instead, they combine the groups in the same sample to achieve a wide range of autistic traits in their sample. Williams et al. (2018), to my knowledge, is the only study that directly tests deception detection in autistic adults. They used the Cheating Task (T. R. Levine, 2007–2011): senders were invited to take part in quiz, and when the examiner stepped out of the room leaving the answer key behind, they were encouraged to cheat by a confederate posing as a fellow quiz-taker. After the quiz, the sender was interviewed about cheating during the quiz. All senders included in the deception detection task said they did not cheat, half of whom were liars and half were truth tellers. Senders were also split in terms of transparency i.e. the degree to which one leaks cues that helps other judge their truthfulness. Half were identified as being transparent – those who had an honest demeanour and told the truth, or those who appeared dishonest and lied, and the other half were non-transparent i.e. those who looked dishonest but told the truth or those who looked honest but lied. (T. R. Levine, Shaw, et al., 2010).

Given people in general are not very adept at detecting deception – only very slightly over chance at 54% accuracy (C. F. Bond & DePaulo, 2008), Levine (2010) credits this to the fact that deception detection is easier for the transparent senders and difficult for the non-transparent senders. Williams et al (2018) tested both autistic and non-autistic adults and found that this difference between detectability of transparent and non-transparent senders was true for both groups, however autistic receivers performed significantly worse than non-autistic receivers in the transparent condition. This suggests that while non-autistic receivers are capable of judging deception from transparent senders – potentially due to their ability to read the senders' intentions and mental states (i.e., mentalize) – but not from non-transparent senders, autistic receivers have been found to struggle even with transparent senders – which could stem from difficulties in mentalizing.

This study, while the only one investigating deception detection in autistic individuals beyond childhood does so with a very limited sample size – only 27 receivers in each group. Additionally, while the Cheating Task used by Williams et al.'s (2018) study simulates real life where senders self-select when lying in a high stakes situation, it is confrontational in nature with the sender responding to multiple follow up questions. As explored in chapter 2, this is not how deception in real life takes place in most scenario – i.e., situations where an individual is interrogated / interviewed after committing a transgression – and deception detection ability measured on a task such as this may not generalize to situations, for example, where a bully lies to victimize an unsuspecting peer.

The present study aims to replicate Williams et al.'s (2018) findings in a bigger sample and build on it by including pre-adolescent and adolescents (11-17-year-olds) along with adults using Levine's (2007–2011) Cheating Task. Additionally, I use the novel deception detection tasks created in Chapter 2 to test if the results generalize to scenarios that are not interrogatory or in conversations that are not in the form of question / answer dialogues. The following are my predictions:

> H_6: Autistic traits will be negatively associated with deception detection ability i.e., receivers reporting higher autistic traits would be worse at detecting deception.
>
> H_7: Autistic individuals would demonstrate difficulties in deception detection in comparison to non-autistic individuals on all tasks. On the Cheating Task, while both autistic and non-autistic individuals will struggle to detect lies from non-transparent senders, autistic senders will also show reduced accuracy for transparent senders.

4.1.5. Impact of poor deception detection: peer-victimization and mental health

As established, deception is ubiquitous in our daily lives, and some instances are harmless – such as white lies – while others are more malicious in nature – such as misleading a peer with the aim to bully them. For individuals who already find social navigation difficult, such as autistic individuals, not being able to tell when someone is being honest and direct with them, could be incapacitating and make them more vulnerable to victimization.

4.1.5.1. Mentalizing, deception detection, and peer-victimization

It was previously thought that bullies lacked the ability to correctly process social information (Randall, 1996), however this view has been challenged in favour of the idea that bullies, especially those who are the ringleaders, must have a keen understanding of others' mental states and social situations to be able to manipulate others while remaining undetected (Sutton et al., 1999a). Sutton et al. (1999b) found that bullies scored much higher on social cognition tasks than victims. In fact, a meta-analysis concluded that there is a positive association between mentalizing and bullying behaviour, showing that bullies have a well-developed understanding of their victims' and others' mental states (Smith, 2017).

It can be thus assumed that bullies use their well-developed mentalizing skills to manipulate others, and that they may use deception as one of the ways they accomplish such manipulations. In fact children and adult males who are better deceivers have also been found to be more domineering, for example children who were judged to be honest when convincing a peer that a foul tasting drink was delicious had a tendency to be dominant and were more likely to be terrorizing in the playground (Keating & Heltman, 1994). Deception has also been found to be a prominent component of verbal bullying (Maboyana & Sekaja, 2015). On the other hand, lower ToM has been associated with higher likelihood of being victimized (Shakoor et al., 2012; Wahyuningsih & Novitasari, 2016), however whether or not a difficulty in detecting deception also predicts vulnerability to victimization is yet to be seen.

It is therefore possible that individuals who may have difficulties mentalizing and who struggle to detect deception, such as autistic individuals, could be exceptionally vulnerable to bullying. Hwang et al. (2018) found that autistic students were more likely to be victims of bullying than bullies themselves, and more likely to be victims of bullying than their non-autistic counterparts (Hwang et al., 2018). Parents of autistic adolescents, teachers, and autistic adolescents themselves report more instances of being the target of peer victimization (Carter, 2009; Little, 2002; van Roekel et al., 2010). In fact, prevalence of peer victimization in autistic individuals is reported to be 45%, compared to 10% in non-autistic individuals (Sterzing et al., 2012).

It has be extrapolated that difficulty in detecting deception can be a possible reason why autistic individuals are the target of peer-victimization more often than their non-autistic peers (Ranick et al., 2013). Sofronoff et al. (2011) interviewed parents to identify areas of concerns that lead to bullying victimization in autistic children and found gullibility and vulnerability to deception to be one of the themes that parents mentioned. This relationship has never been tested directly before, though. This study aims to empirically test this relationship by regressing vulnerability to peer victimization as reported by autistic and non-autistic individuals on their mentalizing and deception detection ability, and it is expected that,

> H_8: Autistic individuals will report higher peer-victimization than non-autistic individuals will.

> H_9: Mentalizing ability and deception detection ability will negatively predict peer-victimization such that better mentalizing and deception detection ability will predict lower levels of victimization reported.

4.1.5.2. Deception detection, peer victimization, and psychological distress

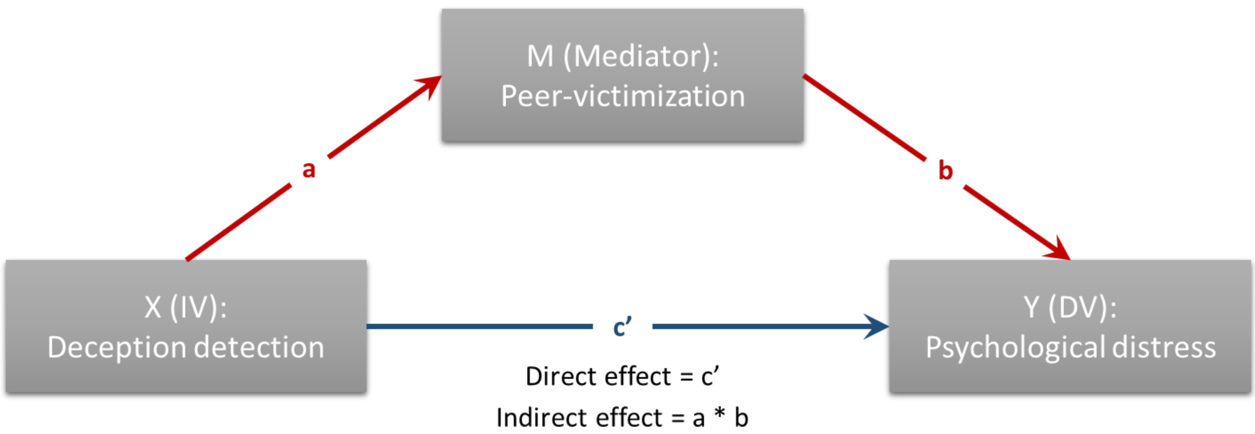
A plethora of research has indicated that the targets of bullying and peer victimization have increased mental health issues including a higher likelihood of having lower self-esteem and being diagnosed with depression and anxiety. This has been found to be true for children (Armitage, 2021; Kumpulainen et al., 2001; Rigby, 2005), adolescents (Chang et al., 2013; Coggan et al., 2003; Ford et al., 2017; Le et al., 2017), and adults (see Verkuil et al., 2015 for review) alike. This is also true for autistic individuals (Cappadocia et al., 2012; Rodriguez et al., 2021; Shtayermman, 2007). Moreover, bullying victimization in autistic samples has also been identified as a significant predictor of suicidality (Holden et al., 2020; Shtayermman, 2007).

While theoretically, there does not appear to be any cause to assume a direct relationship between deception detection ability and mental health, if a difficulty detecting deception contributes to the propensity to being victimized, it is possible that this can further lead to psychological distress and mental health issues both in autistic and non-autistic individuals. If evidence for such an indirect relationship between deception detection and mental health is found, it could mean that interventions based on deception detection training aimed at reducing vulnerability to peer-victimization could potentially help minimize risk of mental health issues or psychological distress. This relationship between deception detection and mental health has, to my knowledge, never been investigated, and this study aims to explore if accuracy on the deception detection tasks predicts the extent of anxiety and depression symptomology reported by participants. Extrapolating from past literature, I propose that,

> H_{10}: Deception detection ability will predict psychological distress, mediated by peer-victimization. In other words, the better the deception detection ability, the lesser will be the reported psychological distress, aided by a lower reported peer-victimization (Figure 4.1)

Figure 4.1

Proposed mediation model of deception detection as a predictor of psychological distress, mediated by peer-victimization



4.2. Method

4.2.1. Participants

4.2.1.1. Adult recruitment and sample

Data was collected from adult autistic and non-autistic participants (18 – 30 years) in two waves that used two different recruitment strategies.

In the first wave of recruitment, participants were invited to participate in the Autism@ICN coordinated testing initiative at the UCL Institute of Cognitive of Neuroscience through advertisements on social media (Twitter, Facebook, Instagram, Reddit). People already on the Autism@ICN database were invited to participate via direct emails. As part of the coordinated testing, participants took part in a background battery and multiple different studies. Participants were paid at the departmental rate of £7.50 / hour (in Amazon UK vouchers or $10 per hour in Amazon US vouchers) for the Autism@ICN background battery they completed. Instead of an hourly rate, on completing the rest of this specific study, participants were entered into a prize draw where one in every 10 participants won an additional £20 Amazon voucher. Participants were required to be fluent in English to participate in the testing session.

Two hundred and twenty-two participants were recruited in this wave; however, data from only 99 participants was included. A set of strict data quality-based exclusion criteria was developed due to a) the online nature the study, b) the open and uncontrolled nature of the recruitment call. This was done to make sure that only participants who were paying attention and putting in effort in the tasks were retained in the final analysis.

The verbal ability task, Spot the Word (STW, Baddeley et al., 1993), and the non-verbal reasoning task, the Matrix Reasoning Item Bank (MaRs-IB, Chierchia et al., 2019), were used to develop this set of exclusion criteria. One hundred and eighteen participants were excluded due meeting one or more of the seven exclusion criteria (Table 4.1) and four were excluded because they reported to be self-identifying autistic individuals but not to have a diagnosis from a clinician.

There was a concern that exclusion criteria disproportionately excluded autistic individuals, however, as most number of participants were excluded on being at or below chance performance or responding to quicker than feasible if attempting the task properly, these were likely fraudulent participants who took part just to earn the compensation and possibly posed as autistic individuals. This was further corroborated by the fact that for a majority of these participants, the email addresses were either repeats of past participants or part of a series of similar email IDs.

A second wave of participants was recruited via Prolific (www.prolfic.co). The decision to recruit from Prolific, despite the additional costs, was made as a response to the high exclusion rate faced in Wave 1 due to poor quality data. Prolific is known to be a research centric participant database with strict screening processes and more transparency (Palan & Schitter, 2018). Participants were all from the UK and fluent in English. To avoid high exclusion rates during analysis (as in Wave 1), the study was set up to reject any participants who scored below 60% on the STW and they therefore did not take part in the rest of the study. The 60% cut-off was devised because the STW was used as both an attention check and as a data quality measure along with its pre-screening for adequate verbal ability – 16-31 year-old non-autistic individuals are on average 77% accurate, the cut-off was lowered considering the younger sample in this study. Of the additional 299 participants recruited in this wave, eight participants were excluded because they met one or more of the criteria mentioned in Table 4.1, leaving usable data from 291 participants. Participants recruited from Prolific were paid at the departmental rate of £7.50 / hour.

In total, data from 390 adult participants was included (M_{age} = 23.94 years, SD_{age} = 3.39 years), of whom 104 were autistic and 286 were non-autistic.

Table 4.1

Exclusion Criteria for Wave 1 of adult recruitment

Measure	Exclusion Criteria	*n* Excluded
MaRs-IB	Completing fewer than 13 items in 8 minutes[1]	2
	At or below chance performance	13
	Responded to more than 10% trials in 250 ms or less[2]	41
	Timed out (did not respond) on 25% or more trials[2]	2
STW	Responded to more than 10% trials in 100 ms or less	30
	Failed more than two of the four practice trial	12
	At or below chance performance	42

Note. The exclusion criteria were devised based on participants' performance on the two tasks used to estimate their verbal ability (STW: Spot the Word, Baddeley et al., 1993; adapted for online testing as seen in Livingston et al., 2021) and non-verbal ability (MaRs-IB: The matrix reasoning item bank, Chierchia et al., 2019). Participants were excluded for meeting one or more of the criteria.

[1]As per Chierchia (2019)

[2]Adapted from exclusion criteria used by Chierchia (2019)

4.2.1.2. Adolescent recruitment and sample

Adolescent participants (11-17 years) were recruited through secondary schools, charities (eg. National Autistic Society), and with the help of participant databases (e.g. Cambridge Autism Research Centre). Participants were required to be fluent in English, resident of the UK, and being educated in mainstream school (or working at an equivalent level).

Similar to the second wave of adult testing, participants were automatically rejected if they scored below 60% in the STW and did not proceed to complete the study. Four hundred and twelve participants completed the study, of whom 14 were excluded for not having a clinical diagnosis (self-identifying or in the diagnostic process), 15 were excluded as they were not recruited through trusted channels, such as schools or charities we advertised through. This was done to minimise the possibility of fraudulent participants who did not meet the age criteria taking part in the study, i.e. to avoid issues faced in Wave 1 of adult recruitment. An additional 33 were excluded for meeting one or more criteria in Table 4.1, and three were excluded due to inconsistent demographic information provided. To compensate for their time and effort, participant were sent £10 amazon vouchers – calculated based on under 90 minutes of testing time at the departmental rate of £7.5 / hour to compensate for their time and effort.

The final adolescent sample consisted of 347 participants (M_{age} = 14.09 years, SD_{age} = 1.82 years), of whom 43 were autistic and 304 were non-autistic.

4.2.1.3. Final Sample

Our final sample consisted of 737 participants (M_{age} = 19.30 years, SD_{age} = 5.64 years), of whom 147 were autistic and 590 were non-autistic. All autistic participants confirmed that they had been diagnosed by a qualified clinician; however, as this was an online study, it was not possible to verify the diagnosis of the autistic participants. Nevertheless, autistic participants reported having significantly more autistic traits, thus substantiating their diagnosis as a group.

The sample size for this study was projected based on past studies. For non-autistic participants I estimated the sample from the study introducing the MaRs-IB task (Chierchia et al., 2019) as they a) tested a similar age range of participants (11 to 33 years), and b) administered the task online. I also took cue from a study looking at development of mentalizing (7 to 27 years) by Dumontheil et al. (2010), and multiplied their sample size by three to account for the fact that the study took place over a decade ago and because unlike their study the current study took place online. Dumontheil et al. (2010) had 177 participants, three times of which was 531, and Chierchia et al (2019) had 659 participants. The final sample estimation for non-autistic group was fixed at 600, half of them adults. Ideally, I would have estimated a similar sample size for the autistic group to make it comparable to the non-autistic group. However, as autistic individuals are harder to reach and recruit, evidenced by underpowered autism studies (Amaral, 2016; Thurm & Swedo, 2012), it was predicted that such a comparable sample size will not be achievable. Therefore, for the autistic group I optimistically estimated 200 participants – 100 each of adults and adolescents – and had a rolling recruitment to realistically recruit as many participants as possible.

4.2.2. Measures

The entire study was designed and hosted on Gorilla Experiment Builder (www.Gorilla.sc).

4.2.2.1. Deception Tasks

The study used three deception tasks – an online adaptation of the cheating tape interviews (T. R. Levine, 2007–2011) and versions of the Weird Picture and Two Truths and a Lie tasks. The battery also included a Demeanour task.

4.2.2.1.1. Cheating task – Levine cheating interview tapes

The Cheating task was an online adaptation of the task used by Williams et al. (2018). Initially created by Levine et al. (2011) using videos shot as part of the National Science Foundation funded cheating interviews (T. R. Levine, 2007–2011), they consisted of videos of people being interviewed after what they assumed was a federally funded university approved teambuilding study, where they played a trivia game with another participant (who was actually a confederate). During the trivia game test session, the examiner left the room for a while, leaving the answer key behind, making it easy for the participant to cheat at the confederate's suggestion – some did and some didn't. Following the test session, Levine's participants were asked a number of general questions followed by three questions about any cheating behaviour that may have taken place while the examiner was absent during the test session. Each question was more pointed than the previous; each intended to put more pressure on the participant to either confess or leak cues of deception:

Q1: Did any cheating occur when the experimenter was out of the room?

Q2: Are you telling me the truth?

Q3: What will your partner say when I ask her the same question?

All lies were unsanctioned and had real consequences if the participant was to get caught – participants were under the impression that by cheating and lying about cheating they were violating a federal funded research study, thus violating university codes which could have serious consequences – making these high stake lies.

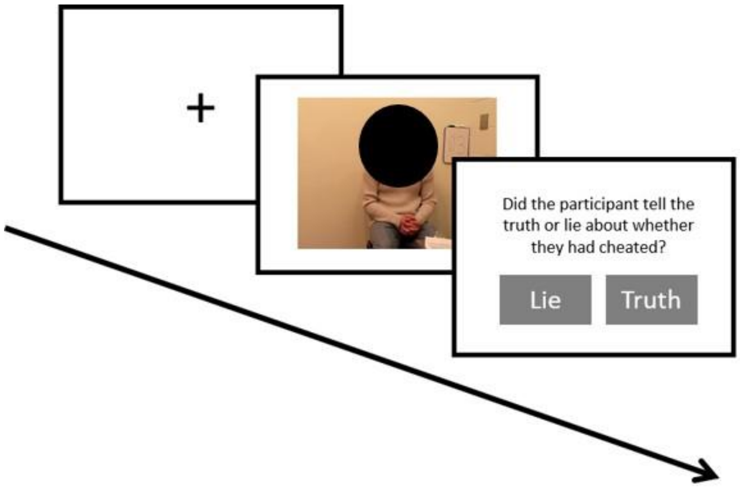
Levine et al. (2010) had a number of receivers judge 44 of these videos – 22 of people who cheated and lied, and 22 matched videos of people who did not cheat and told the truth. Based on these judgements, Levine et al. (2011) finally shortlisted 20 videos: 10 transparent senders, i.e. those whose demeanour and veracity were matched (5 honest appearing truth-tellers and 5 dishonest appearing liars), and 10 non-transparent senders, i.e., those whose demeanour and veracity were not matched (5 dishonest appearing truth-tellers and 5 honest appearing liars).

The cheating interview tapes were adapted for online testing for the current study. Instead of using all 20 videos as experimental trials, two of the videos were used as practice trials. This was a departure from how the tapes are usually administered, and more specifically how Williams et al. (2018) use it in their study. This was done to familiarise the participants with the task because unlike offline lab-based experiments where the experimenter is present to assist the participants, online administration meant that it could not be confirmed that participants understood the set up and instructions.

Therefore, the Cheating Task had two practice trials followed by 18 experimental trials – 10 transparent videos and 8 non-transparent. I decided to use non-transparent videos as practice trials in order to keep the number of videos in the transparent condition the same as Williams et al. (2018), as they found that the difference between autistic and non-autistic participants was most evident in the transparent condition. Each video played only once in a random order. Figure 4.2 represents a trial in the cheating task.

Figure 4.2

Schematic representation of a trial in the Cheating Task



Note. Faces of senders are redacted for the purpose of this report.

Participants had 10 seconds to make their judgement after every video by clicking on the appropriate button on the screen – ‘Truth’ or ‘Lie’. At the end of the task, participants were shown their total score.

The Cheating task took approximately 10 minutes to complete.

4.2.2.1.2. Weird Pictures – Development of Deception Detection study (WP-3D)

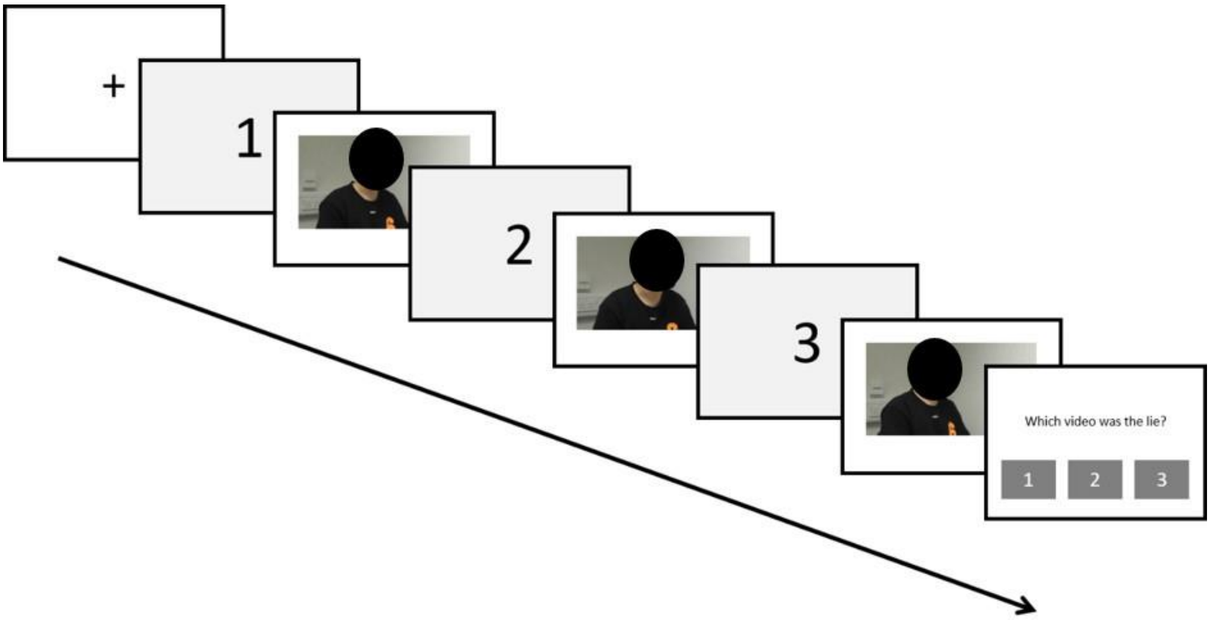
The Weird Pictures task (WP-3D) for this study was created using the WP-final task developed in Chapter 2.

The task started with a practice trial. No feedback was given for this trial unless the participants timed out, in which case they were reminded of the five-second time limit. The actual experiment consisted of four trials presented in a random order. Each trial started with a fixation cross (displayed for 250 ms) followed by of a triplet of videos (one lie, one double bluff, and one truth) with their order randomized within trial; each video was preceded by a screen on which either ‘1’, ‘2’, or ‘3’ was displayed to help participants demarcate the videos.

At the end of the video triplet, a response screen with three buttons with the numbers ‘1’, ‘2’, or ‘3’ on them was displayed and participants had five seconds to respond by clicking on the button they thought corresponded to the lie (Figure 4.3). They had 5 seconds to respond and a timer appeared in the last 3 seconds. If the participants did not respond within the time limit, the task would auto-advance to the next trial. At the end of all four trials, participants’ score and their accuracy percentage was displayed. The WP-3D took approximately 10 minutes to complete.

Figure 4.3

Schematic representation of a trial in the WP-3D



Note. Each trial had a triplet of videos presented in a random order. Before each video a screen with the numbers ‘1’, ‘2’, or ‘3’ was displayed on the screen. At the end of the three videos, a response screen with three response buttons was presented. Faces of senders are redacted for the purpose of this report.

Instructions. A video recording of a researcher informed participants that they were going to watch videos of people describing some weird pictures, which had some unusual things happening in them, to another person off-camera. The participant was shown an example weird picture and then was informed that of the three pictures each person described, for one picture they would describe it truthfully, for one they would double bluff, and one they would lie. Their (the participant's) job was to figure out which one was the lie. They were informed that they would have 5 seconds after watching each person's three descriptions to guess which of them was the lie. The instructions were also reiterated in textual form. At the end of the task, participants were shown their deception detection score on this task.

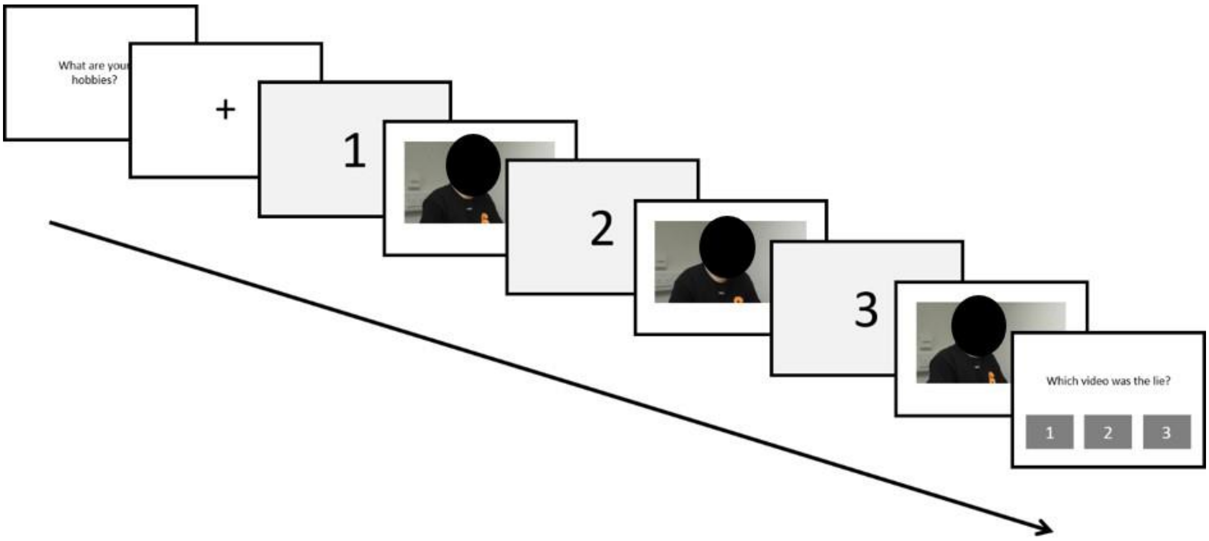
4.2.2.1.3. Two Truth and a Lie – Development of Deception Detection study (TTL-3D)

The Two Truths and a Lie task (TTL-3D) for this study was created using the TTL-final task developed in Chapter 2.

Participants went through a practice trial followed by four experimental trials. Trial order was randomized, as was the video order within each triplet (two true answers and one lie). For every trial, the first screen displayed the question that the specific sender was answering, followed by a fixation cross (250 ms). Then screens with the numbers '1', '2', and '3' were displayed, each followed by one video (one answer). After all three videos was the response screen with three response buttons, one for each of the videos (Figure 4.4). Participants would automatically be taken to the next trial if they did not respond within the time limit. Finally, at the end of the tasks they were shown their scores and accuracy percentage. The TTL-3D took approximately 10 minutes to complete.

Figure 4.4

Schematic representation of a trial on the TTL-3D



Note. Each trial had a triplet of videos presented in a random order. The first screen displayed the trial-specific question. Before each video, a screen with the numbers ‘1’, ‘2’, or ‘3’ was displayed on the screen. At the end of the three videos, a response screen with three response buttons was presented. Faces of senders are redacted for the purpose of this report.

Instructions. As with WP-3D, TTL-3D started with video instructions, followed by textual instruction. Participants were informed that they would watch videos of people playing a game called “Two Truths and a Lie” with another person where they answer questions about themselves, and for every question they gave three answers, two of which were true and one of them was a lie. Their goal was to guess which one of the three answers was the lie. As with the WP-3D, participants were instructed on how to respond (selecting one of the buttons corresponding to the video number) and how much time they had after every triplet (5 seconds). As with the WP-3D, participants saw their score at the end of the task.

An additional disclaimer informed participants that the people in the videos played this game before the Covid-19 pandemic. Therefore, their answers were from before the time when meeting friends and travelling was made difficult, and the participants should keep this mind when judging the videos.

4.2.2.1.4. Demeanour task – Development of Deception Detection study (DT-3D)

The Demeanour task (DT-3D) for this study was created using the Demeanour task used in Pilot 2, as described in Chapter 2.

The task consisted of silent videos of all the senders appearing in either WP-3D or TTL-3D. Participants had to rate how honest the senders appeared from 1 (most dishonest) to 7 (most honest). Each trial consisted of one video, 20 seconds long. Participants had 5 seconds to respond after watching the video. As there were four senders in each of the WP-3D and the TTL-3D, there were eight experimental trials in the DT-3D, plus a practice trial. The DT-3D took about 5 minutes to complete and was always administered before the WP-3D and TTL-3D.

4.2.2.2. Mentalizing task: Frith-Happé Animations (AT)

Mentalizing ability was measured using the Frith-Happé Animations (Abell et al., 2000), specifically, the objective-scoring version created by White et al. (2011), described in detail in Chapter 2. The task was adapted for online testing by Livingston et al. (2021).

Although based on White et al.'s (2011) short (objective) version, Livingston et al. (2021) did not include the multiple choice questions (MCQ-Feelings) that participants were asked in White et al.'s (2011) version if they categorised the Theory of Mind (ToM) animations correctly. For this study, some further changes were made to the Livingston et al.'s (2021) online version in order to reduce the total time spent on the task: since only the objective scoring was used, participants were not required to type a subjective description of the animation after categorizing it.

Other modifications to the online version were made after feedback from pilot participants about the ambiguity of the instructions and descriptions. Instructions were also revised to make the task instructions more accessible to children, especially considering there were no experimenters present to clarify any confusions and give feedback after practice trials. Table 4.2 details the changes to the descriptions of the three animation categories that we made.

The final online task used in this study had three practice trials, one of each type (Random, General Direction, and ToM) with feedback, followed by 12 experimental trials, randomized and without any feedback. After every animation played to completion, participants categorised the animation by clicking one of the three buttons on the response screen, which had the full descriptions (minus the examples). They had six seconds to make their response and a timer counted down the last five seconds, informing participants of the time limit. The task took, on an average, about 12 minutes to complete.

(Rest of page intentionally left blank)

Table 4.2

A side by side comparison between the AT descriptions as developed by White et al. (2011) and the modifications made for this study

Animation type	Objective Categorisation	Description	
		White et al. (2011)*; Livingston et al (2021)	This study
Random	"No Interaction"	"There is no obvious interaction between the triangles and movement appears random."	"There is no obvious interaction between the triangles - the movement appears random."
General Direction (GD)	"Physical interaction"	"An interaction between the triangles in which actions are directed toward each other in order to achieve specific goals."	"There is an interaction between the triangles - they are both acting for a specific purpose, such as playing a simple game."
Theory of Mind (ToM)	"Mental interaction"	"An interaction between the triangles involving the manipulation of the emotions and thoughts of one triangle by the other."	"There is an interaction between the triangles - one triangle manipulates the thoughts or feelings of the other, such as playing a trick on someone."

*Note.** Instructions / descriptions taken verbatim from White et al. (2011)

4.2.2.3. Questionnaire measures

4.2.2.3.1. Autistic traits: Autism Spectrum Quotient (AQ)

The Autism Spectrum Quotient (AQ, Baron-Cohen et al., 2001) was used to measure autistic traits. The AQ is a 50-item self-report questionnaire that records participants' responses on a 4-point scale ('slightly agree', 'definitely agree', 'slightly disagree', and 'definitely disagree'). An online version of the questionnaire, available in Gorilla's resource database (www.gorilla.sc), was used for this study.

While the AQ is primarily for the adult population, it does have an adolescent version that would have been more appropriate for our adolescent participants. However, in the spirit of keeping measures the same throughout our sample, the adult questionnaire was used after making changes to the language of some items and/or adding definitions of difficult to understand words in order to make it more accessible to the younger participants. To ascertain necessary modifications, the adult AQ was attempted by four 12-18 year-old fluent English speaking volunteers, who gave feedback on the comprehensibility of the AQ. Table 4.3 shows the three necessary changes / additions that were identified.

Table 4.3

A side by side comparison between the AQ descriptions as developed by Baron-Cohen et al. (2001) and the modifications made for this study

Original item (Baron-Cohen, Wheelwright, Skinner, et al., 2001)*	Modification
"I tend to have very strong interests, which I get upset about if I can't pursue."	"I tend to have very strong interests, which I get upset about if I can't explore or engage in."
"I enjoy doing things spontaneously."	"I enjoy doing things spontaneously (without planning for them beforehand)."
"I am a good diplomat."	"I am a good diplomat (a diplomat is a person who can deal with others in a sensitive and tactful way)."

Note. *Item descriptions taken verbatim from Baron-Cohen et al. (2001)

4.2.2.3.2. Bullying victimization measure – the Multidimensional Peer Victimization Scale – Revised (MPVS-R)

The Multidimensional Peer Victimization Scale – Revised (MPVS-R, Betts et al., 2015) was employed to record and quantify the prevalence of peer victimization in our sample. The MPVS-R is a 20-item self-report measure that records how frequently participants experienced each event in the last school year on a three-point Likert-type scale, where 1 = "not at all", 2 = "once", 3 = "more than once".

Questions were divided in five domains:

- Physical (e.g.; "how often has someone kicked you?"),
- Social manipulation (e.g., "how often has someone tried to turn your friends against you?")

- Verbal (e.g., “how often has someone made fun of you for some reason?”)
- Attack on property (e.g., “how often has someone stolen something from you?”)
- Electronic (e.g., “how often has someone sent you a nasty text?”)

The central instruction of the questionnaire, “how often during the last school year has another pupil done these things to you?”, was modified to be appropriate for children and adults alike: “Please answer these questions about your experience in the last 2 years.” The time range for the experiences was increased from approximately one year (one school year) to two years as it was assumed that, even with restriction having been lifted during testing (Spring 2022), the frequency of participants’ one to one interactions were reduced during the previous years due to the Covid-19 pandemic.

The scale was scored such that the higher a participant scores, the higher their experience of peer-victimization. The only study reporting internal consistency of the MPVS-R used a Chinese version of the scale, but they did report a high Cronbach’s α (0.81; L. Li et al., 2020). The Multidimensional Peer Victimization Scale (MPVS), an earlier version of the scale, which only differs from the MPVS-R in that it does not have the ‘electronic’ sub-scale, has been seen to have high internal consistency and acceptable split half reliability (see: S. Joseph & Stockton, 2018). Although the scale was created for children and adolescents, it has been used in adult samples as well (S. Joseph & Stockton, 2018). While it has not been validated for autistic individuals specifically, past studies have used the scale to measure victimization in autistic samples (Brown et al., 2019).

4.2.2.3.3. Mental health assessment – the Hospital Anxiety and Depression Scale (HADS)

The Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983) is a short 14-item scale measuring anxiety and depression on a 4-point scale. It is divided into two subscales, HADS-A quantifying anxiety and HADS-D measuring depression, with 7 items each. Items are scored from 0 to 3, with some items being reverse scored. Higher scores indicate higher symptomatology, with a suggested clinical cut-off of 8 for each subscale.

In a systematic review of 747 papers that used the HADS, Bjelland et al. (2002) found that it had high mean Cronbach's α for both subscales (HADS-A: 0.83; HADS-D: 0.82), and had good to very good concurrent validity when compared to other measures testing anxiety and depression. The HADS has also been validated in various different cultures and on a number of non-English speaking populations (E.g.: Muszbek et al., 2006; Michopoulos et al., 2008; Risal et al., 2015; Al Aseri et al., 2015; also see: Herrmann, 1997).

4.2.2.4. Background measures

4.2.2.4.1. Verbal ability estimate – Spot the Word (STW)

The Spot the Word task (STW; Baddeley et al., 1993; adapted for online testing by Livingston et al., 2021) is a word categorisation task that was used to obtain an estimate of verbal ability. Participants are required to select the actual word in a pair of words comprising of a word (e.g. oasis) and a non-word (e.g. broxic). The task consists of 60 pairs of words as part of the experimental block, with four additional pairs that make up the practice trials to familiarise participants with the task. Participants are instructed to respond as quickly and accurately as they can, though unbeknownst to them, there is no time limit on each trial. A higher accuracy percentage can be extrapolated to mean better verbal ability.

Online administration of the battery made use of traditional verbal reasoning tasks such as those within the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2008) impossible. The STW, however, has demonstrated convergent validity to the WAIS (Yuspeh & Vanderploeg, 2000), and therefore it was deemed as an appropriate proxy measure for verbal ability.

4.2.2.4.2. Non-verbal reasoning ability estimate – Matrix Reasoning Item Bank (MaRs-IB)

The Matrix Reasoning Item Bank (MaRs-IB; Chierchia et al., 2019) is a matrix based non-verbal reasoning task that was used to extrapolate performance ability. It has 80 incomplete 3 x 3 matrices with eight of the cells filled with abstract shapes. Participants had to choose the shape that completed the given matrix best, by evaluating the relationship between the other shapes in the matrix. The task was set up such that participants had 30 seconds to respond in each trial (with a countdown appearing in the last 5 seconds), although they were told to respond as fast and as accurately as they could.

The first block was a practice block where participants had to get three correct to advance to the experimental block. The experimental block ran for 8 minutes and all 80 items were presented in the same order for all participants, starting with the five easiest matrices. Participants were not made aware of the total number of items and neither were they expected to go through all 80 items in 8 minutes. Any participants who completed all 80 items within the 8-minute period continued to redo the items in the same order but the responses were not collected or analysed any further. Feedback in the form of a tick or cross was displayed on the screen after they have made the response, both in the practice and experimental trials.

Scoring of the matrices was calculated as proportions of the number of items attempted rather than as a proportion to the total number of items. A higher score indicated higher non-verbal reasoning ability. The MaRs-IB has been found to have acceptable convergent validity, test-retest reliability, and internal consistency (Chierchia et al., 2019).

4.2.2.5. Technical checks – bot and sound checks

As many of the tasks in the study consisted of audio stimuli, it was essential that participants had functioning audio output devices. We used a Pre-experiment sound-check (James, 2019) that required participants to play an audio file (containing the word "octopus") and type out the word they heard – this tested if the participant could hear the audio and encouraged them to adjust volume levels accordingly. If correctly entered, the task then tested if their device's auto-play functionality was enabled by automatically playing the word "chocolate" which the participants then had to type in a response box. Three consecutive failures to enter either of the two words correctly resulted in automatic rejection.

The sound-check had a second purpose: since bots struggle with free text-response boxes, this task had proven to be a good bot-screening task during pilot testing. Bots and fraudulent data are a huge concern with online studies (Pozzar et al., 2020). Considering that recruitment for this study was open and uncontrolled in nature in Wave 1, i.e., the link was advertised on social media and anyone with the link could participate, we ran the risk of bots populating our study. Therefore, the sound-check assisted us in preventing automated bots from completing the study.

4.2.3. Procedure

Adult participants (18- 30 years) in Wave 1 and Wave 2, and adolescent participants (11-17 years) took part in all the same tasks but the order in which the tasks / questionnaires were presented differed. The whole study was created and hosted on Gorilla Experiment Builder (www.gorilla.sc).

4.2.3.1. Adult Participants Wave 1

The chronology of the tasks / questionnaires in Wave 1 in presented in Figure 4.5. Adult participants in Wave 1 took part in this study as part of a wider coordinated testing project, which comprised of multiple studies / stages. The entire coordinated testing session ran for a maximum of three hours and participants were encouraged to take breaks between stages but complete the whole session within one single day.

Wave 1 participants did all the tasks that were common to all studies in the coordinated testing study as part of Stage 1 (Autism@ICN background battery) and the deception tasks and questionnaires specific to this study as Stage 2. After that, they continued on to other tasks that are not included as part of this study / thesis. Participants started the coordinated testing session by clicking on a direct link that took them to the online testing environment. They were first screened for eligibility after which they read the overall information sheet and filled out a consent form, and finally created unique participant codes to help us link their demographic data to their experimental data in the various stages and studies.

Stage 1 and Stage 2 each took approximately 45 minutes, therefore a total of 90 minutes were spent completing tasks for this study.

4.2.3.2. Adult Participants Wave 2

Adults in Wave 2 were recruited specifically for this study. They joined the study by signing up on Prolific (www.prolific.co). The study was advertised as a multi-stage study, where participants had to finish one stage to be invited into each subsequent stage. There were three stages and each stage was about 30 minutes long. Once in the study environment participants read the information sheet and filled the consent form after an eligibility-screening questionnaire. The chronology of the tasks / questionnaires in Wave 2 is presented in Figure 4.5. Each break denotes the end of a stage.

4.2.3.3. Adolescent Participants

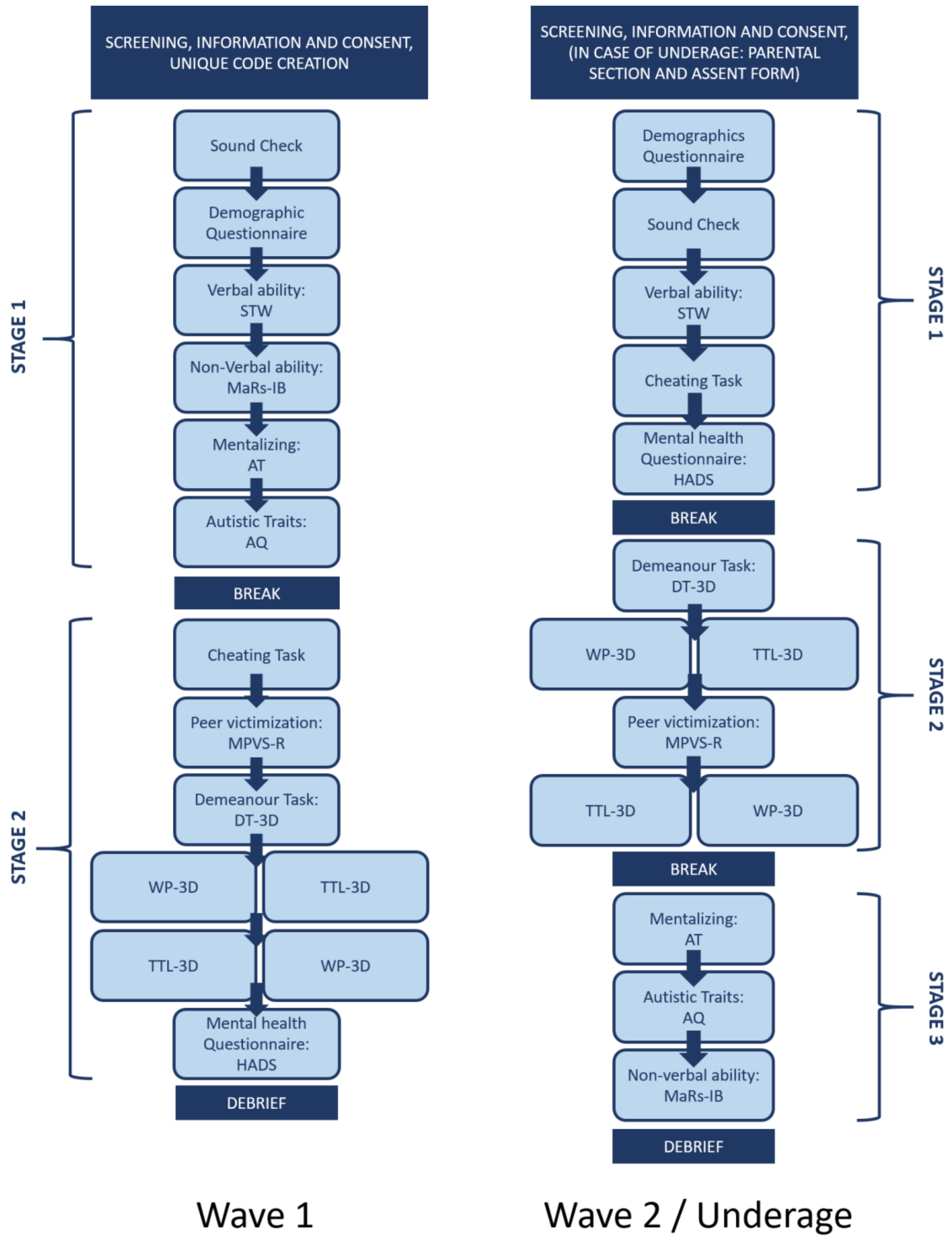
Chronology of the tasks / questionnaires for adolescent participants is presented in Figure 4.5.

Adolescent participants joined the study by clicking a link on the advertisements that took them to the study webpage. Here, parents / guardians had the opportunity find out more about the study and use a link to the study environment on Gorilla (www.gorilla.sc) where they could then start the study if they chose to enrol their children. The study environment was divided into the parental introductory section and the actual study. On entering the study environment, parents found themselves in the parental introductory part of the study, where they were required to read an information sheet and fill the consent form. Parents of younger participants (11-15 years) also created a unique participant code for their children to use throughout the study and filled in demographic information before handing over the study to their children. Parents of older children (16-17 years) handed over to their children immediately after giving their consent, where the participants themselves created their own unique codes. All adolescent participants read an age-appropriate information sheet and filled in an assent form before moving onto the actual study. From there, participants went into the actual experiment, which was divided in three stages, each about 30 minutes long. Participants were encouraged to take breaks between stages.

While participants who were aged 16-17 years filled their own demographic information as part of the actual study, younger children (11-15 years) were only required to refill a brief demographic form (just gender, age and diagnosis) as part of the actual study. While this information was redundant, this was done to make sure that all participants experienced the same number of tasks / questionnaires.

Figure 4.5

Schematic representation of the experimental procedure in Wave 1 and Wave 2 of adult studies, and the adolescent study



Note. The study was divided into multiple stages for both Wave 1 and Wave 2 / adolescent participants and participants were encouraged to take a break between stages in order to reduce fatigue. The order of WP-3D and TTL-3D was counterbalanced between participants so the time lapsed after the demeanour task did not adversely influence the effect of familiarity (induced by the demeanour task) on the performance on only one of the two deception tasks.

4.2.4. Improving participant engagement

Running studies online comes with a host of issues, especially considering that the study involved testing adolescent participants. Various steps were taken to mitigate these issues and to improve engagement:

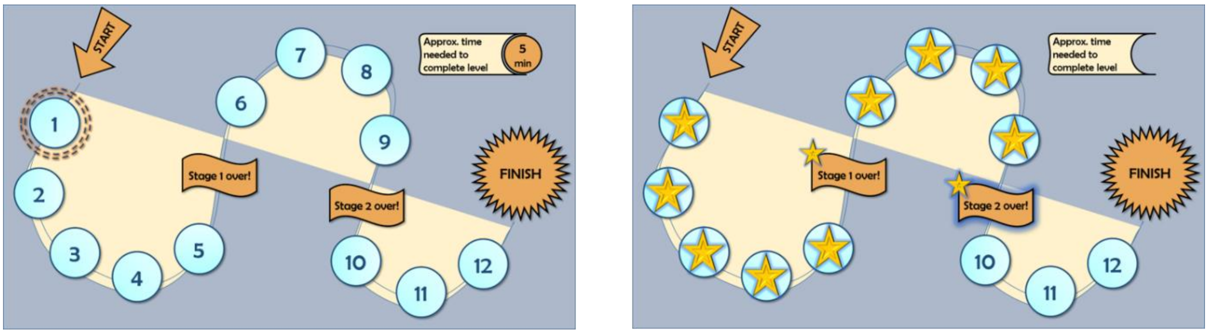
Surroundings check. Participants were required to meet certain criteria to ensure that they were taking part in the study under optimum conditions. They were required to confirm that they had fast internet that would allow for the videos in the tasks to stream immediately. Participants were required to sit comfortably and in a quiet place. It was also recommended that they wear earphones / headphones. Finally, participants were instructed to put their phones on silent to avoid distraction. These surround checks were put in place at the start of every stage.

Engaging instructions. Instructions for the deception tasks were presented via a video recording by a young and engaging researcher at the beginning of the task. Video instructions, especially those delivered in a conversational style have been seen to be helpful in improving participant engagement (Gorilla Support, 2022) and this was especially important as the task was being administered entirely online to adults, and more importantly, to children.

Progress map. It is recommended that participants be given an overview of the study to help them to not feel lost, especially in long multi-task studies (Gorilla Support, 2022). To this end, a map akin to those found in computer games, detailing the various 'levels' (tasks) and stages was created. Participants could track their progress, earning a star every time they completed a task. The map also showed how long the next task would take. It was hoped that it helped them plan their breaks and motivated them to finish subsequent tasks. See Figure 4.6 for some examples.

Figure 4.6

Progress maps to increase participants engagement



Note. Examples of progress maps from Wave 2 / adolescent participants: showing the 'level' (task) participants were about to start with how much time was required to complete the level in the top right corner (left); showing participants their progress through the study – helping them keep track by earning a star for each 'level' (task) they completed (right).

Incentives. To improve motivation and thus performance in a relatively uncontrolled online testing environment, bonus prizes for the 'best lie detector' were introduced for each group of participants (adolescent autistic, adolescent non-autistic, adult autistic, and adult non-autistic). Participants were told that the best performing participant would be eligible to get an additional Amazon voucher worth £20 and therefore they should try their best to detect the lies. They were assured this was over and above compensation for time, and in the case of Wave 1 adults, this was in addition to being entered into the prize draw.

Scores. Participants could find out how well they did in the deception tasks as well as the verbal and non-verbal ability tasks. This was combined with encouraging statements to help motivate the participants even further.

4.2.5. Ethics

All participants (and parents of adolescent participants) read through detailed information sheets detailing all risks and benefits of the study, and in the case of adolescent participants, an assent sheet that explained the risks and benefits in age-appropriate language. The assent sheet was pilot tested by one adolescent autistic and four adolescent non-autistic volunteers to check for comprehensibility. Participants were told that they were free to close the study at any point if they felt distressed without any repercussions.

Some of the tasks / questionnaires in the study dealt with sensitive topics such a peer victimization / bullying victimization and mental health. As such, it was a concern that a remote online study would put participants at risk as there was a possibility that they would be distressed and there would be no way for the experimenter to know. This was mitigated by sign-posting participants to various charities and helplines that could help them with any concerns they may have, or if they felt upset during or after the study. All participants were encouraged to talk to someone they were close to like a close friend or family member (or a trusted adult, in the case of adolescent participants) to further discuss any distress caused by the study. Additionally, parents of adolescent participants were made aware of the risks and were asked to open a conversation with their children after the study. Parents were also given a resource sheet contained information on how to speak to kids about sensitive issues such as those covered in this study, along with details of charities and helplines that specifically deal with children.

Participants who completed the study but who were later excluded due to not meeting data quality standards were paid – this was especially relevant for Wave 1 participants. In Wave 2 and for adolescent participants, participants who were auto-rejected due to underperforming on STW were not paid; however, participants were explicitly forewarned about this in the information sheet and assent forms.

All procedures were approved by UCL Research Ethics Committee (Project ID Number: 14807/002).

4.3. Results

4.3.1. Group matching

The number of the non-autistic participants (NT) in the final sample was much larger (n = 590) than autistic participants (ASC, n = 147). The two groups were matched on verbal ability but not on age (ASC was older), non-verbal ability (ASC scored lower), and gender (ASC had fewer females than males). To make the groups comparable, participants were systematically removed from the sample: NT participants with the highest verbal and non-verbal abilities were removed from the adolescent (females only) and adult samples separately, so as to not exclude the younger participants disproportionally. The exclusions were solely based on demographic and background measures and experimental data was not considered during this process. Along with being matched on gender, verbal and non-verbal ability, this also led to the groups being matched on age within the adolescent and adult samples, so no further removal based on age was required.

As mismatched sample sizes have implications on over all power, type-I error rates, and how variance within the samples are tested (Parra-Frutos, 2013; Rusticus & Lovato, 2019), an equal number of NT participants as ASC participants were selected randomly from both the adolescent and adults samples. The selection was generated in SPSS using the “Select cases → Random sample of cases” function without input of any further variables. This created a NT subsample that was matched with ASC group on age and verbal ability, non-verbal ability, but not autistic traits, which was expected. Groups were also comparable on the gender make-up. For any tests that compared the two groups directly, the ASC group was compared to the matched NT sample. Sample statistics for the ASC group and both the full NT sample and matched NT subsample are in Table 4.4, along with statistics regarding group matching. Figure 4.7 illustrates the process of group matching.

Table 4.4.

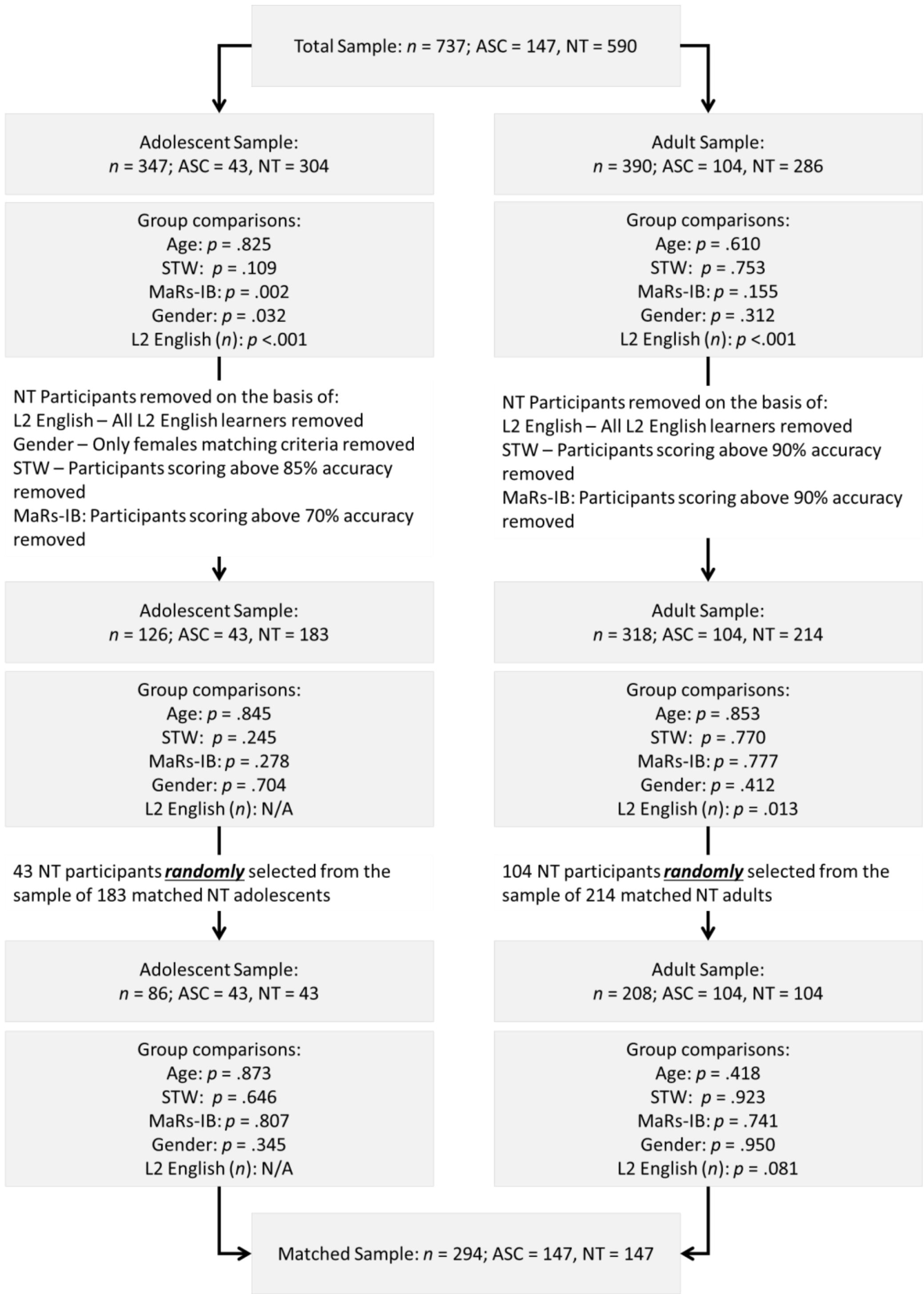
Sample demographics: Autistic sample compared to full non-autistic sample and matched non-autistic sample

	ASC (*n* = 147)		NT (*n* = 590)			Matched NT (*n* = 147)		
	Mean (*SD*)	Range	Mean (*SD*)	Range	*p*-value	Mean (*SD*)	Range	*p*-value
Age (years)	20.94 (5.50)	11-30	18.90 (5.61)	11-30	<.001	21.19 (5.50)	11-30	.695
STW	77.47 (11.49)	51.67-100	78.68 (9.92)	48.30-100	.242	77.84 (9.02)	53.33-98.33	.756
MaRs-IB	55.73 (19.23)	25.64-100	62.50 (18.52)	25.32-100	<.001	56.61 (16.28)	26.25-100	.675
AQ	30.04 (8.94)	13-48	20.04 (6.71)	4-45	<.001	19.86 (7.07)	4-38	<.001
Gender	63 F; 77 M; 7 NB/O		312 F; 261 M; 17 NB/O		.032	66 F; 73 M; 8 NB/O		.476
L2 English learners	3		81		<.001	0		.082

Note. STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), NB/O = Non-binary/Other/Prefer not to say; L2 English learners = Participants who are second language speakers of English

Figure 4.7

Flowchart demonstrating the process of matching the ASC and NT groups on the basis of demographic information (age, gender, English fluency), background measures (verbal and non-verbal abilities), and sample size.



Note. STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), L2 English learners = Second language speakers of English. The combination of exclusion criterion percentages for each of the background measures was determined by trial and error, till groups were matched.

4.3.2. Internal consistency of questionnaire measures

The questionnaires used here were amended in order to either make them more accessible to the younger participant pool, such as in the case of the AQ, or to account for the unusual circumstances created by the pandemic, such as in the case of the MPVS-R and the HADS. Keeping this in mind, the internal consistency reliability of the three questionnaire measures used were examined, determined by the Cronbach's alpha – for the whole sample and also for the groups separately. Cronbach's alpha ranges from 0 to 1, where a higher value indicates higher internal consistency, with 0.70 being deemed as an acceptable level for questionnaires being used in social science research (Kline, 2015).

4.3.2.1. Autism Spectrum Quotient (AQ)

The AQ is made up of five subsets, and the internal consistency of each of the subsets as well as the overall internal consistency of the questionnaire was investigated. The AQ generally has very good overall internal consistency, with alpha values ranging from .67 (Hurst et al., 2007), which is marginal, to .82 (Austin, 2005), which is considered to be very good. The AQ in this study showed very good reliability, surpassing that of even Austin (2005), with a Cronbach's alpha of .86 for the whole sample, .88 for the autistic group, and a more modest but still good .78 for the non-autistic group.

When investigating individual subscales, some were found to have lower Cronbach's alpha, but these lower alpha values were either comparable to or higher than what is found in the literature. The Cronbach's alpha values for all subscales along with those found by Hurst et al. (2007) are provided in Table 4.5.

Table 4.5

Cronbach's alpha values for all AQ subscales

Subscale	Across groups	ASC	NT	Hurst et al. (2007)
n	737	147	590	1005
Social	.77	.80	.72	.66
Attention Switching	.62	.70	.55	.41
Attention to detail	.59	.57	.59	.60
Communication	.72	.69	.63	.47
Imagination	.57	.66	.37	.40

Note. Cronbach's alpha values from Hurst et al. (2007) provided for comparison

4.3.2.2. The Multidimensional Peer Victimization Scale – Revised (MPVS-R)

The MPVS-R showed excellent internal constancy across groups (α = .94), higher than the only known previous study (L. Li et al., 2020) that reports its Cronbach's alpha (α = .81). The ASC group had an α of .96, while the NT group had an α of .93. The scale is made up of five subsets, all of which showed high internal consistency (Table 4.6), both across groups and separately for either group.

Table 4.6

Cronbach's alpha values for all MPVS-R subscales

Subscale	Across groups	ASC	NT
n	737	147	590
Physical	.85	.93	.82
Social	.87	.89	.86
Verbal	.84	.86	.83
Property	.86	.93	.82
Electronic	.88	.92	.86

4.3.2.3. The Hospital Anxiety and Depression Scale (HADS)

The HADS has two subscales, both of which measure different constructs: anxiety (HADS-A) and depression (HADS-D). Therefore Cornbach's alpha for this measure is reported separately for the two subscales in the literature (Bjelland et al., 2002 HADS-A: 0.83; HADS-D: 0.82). In this study, HADS-A was found to have acceptable consistency across groups (α = .78), as well as for ASC group (α = .73) and NT group (α = .78) separately. Similarly, HADS-D too showed good internal consistency across groups (α = .78), for the ASC group (α = .80), and for the NT group (α = .75).

4.3.3. Mentalizing ability

Mentalizing ability was measures by the Frith-Happé animations (AT). The total AT was scored out of 12 and each of the types / conditions (Random, GD, ToM) was scored out of 4. A 3 x 2 mixed Analysis of Variance (ANOVA) was run with animation type (Random, GD, ToM) as the within subjects variable and diagnostic group (ASC / NT) as the between subjects factor (Figure 4.8). A marginally significant main effect of group was detected, F (1,292) = 3.763, p = .053, η_p^2 = .013. NT participants scored slightly higher on average than ASC participants across all animation types. A main effect for animation types was also found, F (2,584) = 14.152, p < .001, η_p^2 = .046, and post-hoc t-tests showed that all conditions were significantly different from each other (Random/GD: p < .001; Random/ToM: p = .003; GD/ToM: p = .037), with participants scoring highest in the AT-Random and lowest in the AT-GD condition. The interaction between animation type and group was also significant (F [2,584] = 5.653, p = .004, η_p^2 = .046); ASC participants scored significantly lower than NT participants on the Random animations (t[273.69] = -3.756, p < .001), but both groups were comparable on the GD animations (t[292] = -1.024, p = .153) and ToM animations (t[292] = 0.053, p = .479). See Table 4.7 for all means and *SD*s.

Table 4.7

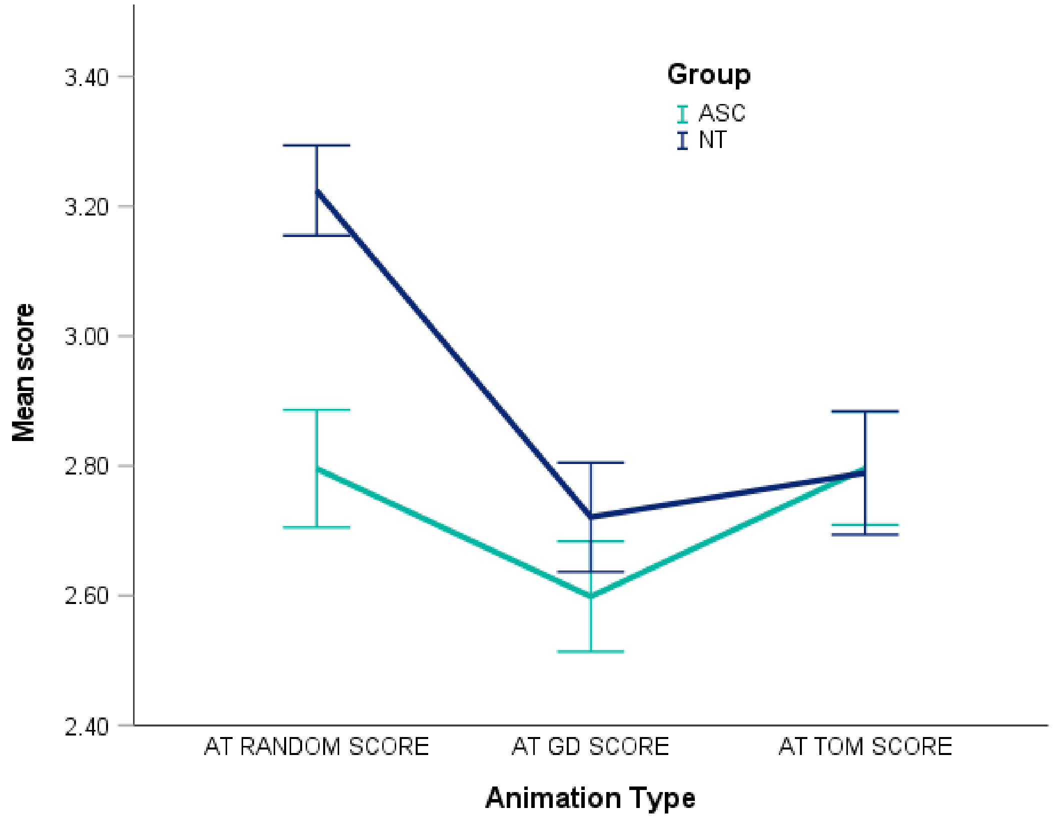
Descriptive statistics for the AT

	Across Groups		ASC (*n* = 147)		NT (*n* = 147)	
	M	*SD*	*M*	*SD*	*M*	*SD*
AT-Random	3.01	1.00	2.80	1.10	3.22	0.84
AT-GD	2.66	1.03	2.60	1.03	2.72	1.02
At-ToM	2.79	1.10	2.80	1.05	2.79	1.15
AT-Total	8.46	2.42	8.19	2.53	8.73	2.28

Note. n = sample size, *M* = mean, *SD* = standard deviation, AT = Frith-Happé animations, GD = Goal directed, ToM = Mentalizing

Figure 4.8

Graphical illustration of participants' AT score



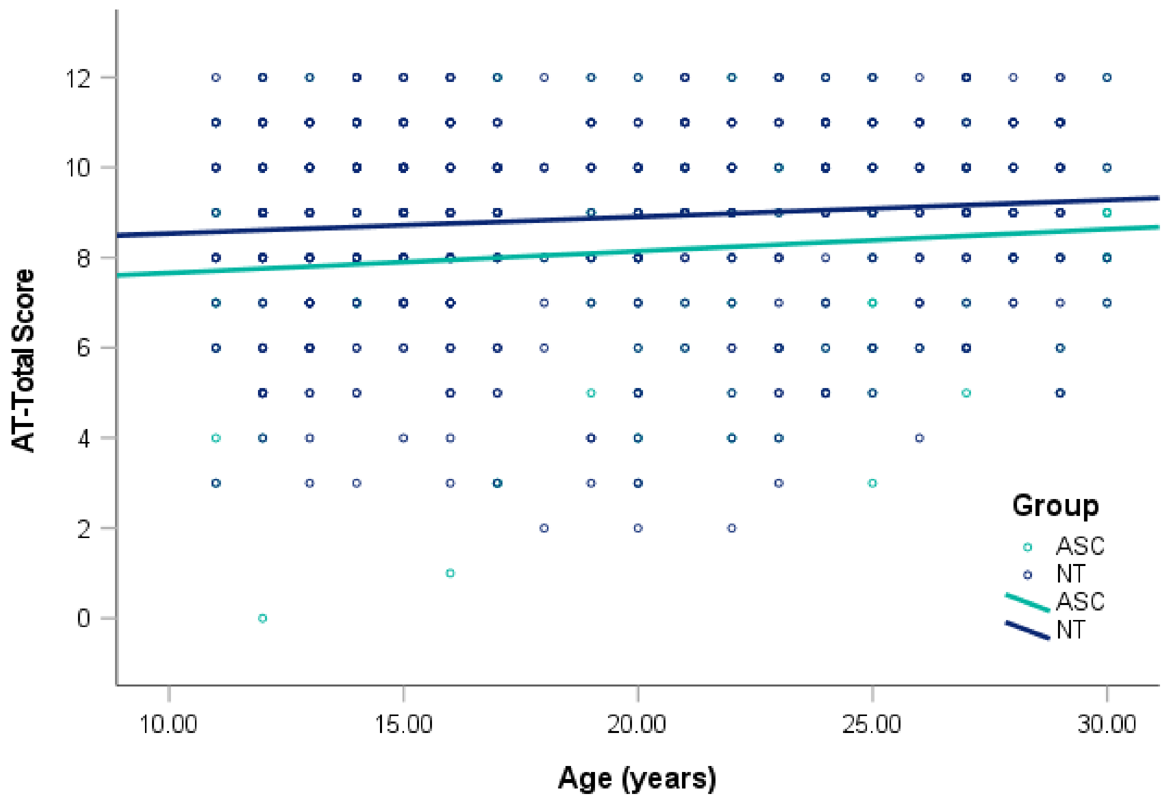
Note. Error bars represent one standard error.

A pair of Spearman's correlations was run to investigate the relationships of age with AT-Total scores (Figure 4.9) and AT-ToM scores, separately for each receiver group. Age correlated significantly with the AT-Total score for NT (r = .102, p = .013), but did not for ASC (r = .070, p = .398). Age did not correlate with AT-ToM for either NT (r = .036, p = .388), or ASC (r = .023, p = .785).

A multiple linear regression was run to see if age predicted the AT-Total score differently in the two diagnostic groups. The interaction between mean centred age and diagnosis was not found to predict AT-Total score; β = -.080, t (289) = -.983, p = .326. A separate multiple linear regression for AT-ToM showed that the interaction was also non-significant for AT-ToM, β = -.044, t (289) = -.535, p = .593.

Figure 4.9

Relationship between AT-Total and age for the autistic and non-autistic group



4.3.4. Deception detection ability

Deception detection ability was measured using three tasks – the Cheating Task (T. R. Levine, 2007–2011) and the novel tasks created for this project – the Weird Pictures (WP-3D) and the Two Truths and a Lie (TTL-3D).

In the Cheating Task, there were a total of 18 trials – eight non-transparent and ten transparent trials. Receivers were scored 1 for every trial they correctly responded to and 0 for every incorrect response. Instead of a total score, accuracy was computed by calculating a corrected hit rate (CHR) for each of the transparency conditions and for the task overall (D. M. Williams et al., 2018):

CHR = [(proportion of truths correctly identified + proportion of lies correctly identified) - (proportion of truths incorrectly judged as lies + proportion of lies incorrectly judged as truths)]

For the WP-3D and TTL-3D, there were four trials each. Participants were scored 1 if they identified the lie (out of the three videos) correctly, and 0 if they did not. To calculate accuracy, scores across trials for each task were summed for each participant. As both WP-3D and TTL-3D had possible accuracy scores of only 0 to 4 and a pair of Shapiro-Wilk tests demonstrated that the accuracy scores were not normally distributed, all analyses were run using non-parametric tests.

Three NT participants were excluded from the analysis for TTL-3D and one ASC participant was excluded from WP-3D due to incomplete data (technical issues).

4.3.4.1. Testing performance against chance

4.3.4.1.1. Cheating Task

A participant with a CHR of zero would be considered to be performing at chance-level. To judge if participants performed above chance, multiple one-sample t-tests were run against a test value of 0 with a more conservative alpha level of 0.01.

Receivers showed above chance performance on the Cheating Task overall. When split by transparency, performance was significantly above chance in the transparent condition, but significantly below chance in the non-transparent condition. Similar results were found for each diagnostic group separately as well. All results were significant at p = .001 level and showed medium to large effects sizes (Table 4.8).

4.3.4.1.2. Novel Tasks

As this was the first time the novel tasks were being used as part of an experimental paradigm, it was important to check that the task was working as expected, to be able to draw conclusive inferences. To do so, firstly, like the Pilot-2 in Chapter 2, it was assumed that for the task to be deemed to be working, non-autistic (NT) participants would need to score above chance.

Analysis similar to Pilot 2 (Chapter 2) was undertaken: one-sample Wilcoxon Signed Rank tests were performed against a test value of 1.33 (chance: 4 trials x 1/3 probability of getting the correct answer by chance). NT participants scored significantly above chance for both WP-3D and TTL-3D (Table 4.9).

Table 4.8

Descriptive and inferential statistics for participants' CHR against chance in the Cheating Task

Group	Condition	*n*	*M*	*SD*	*t*	*df*
Across groups	Total	737	0.33	0.46	19.679	737
	Transparent	737	0.86	0.59	39.884	737
	Non-transparent	737	-0.33	0.76	-12.015	737
Autistic	Total	147	0.26	0.45	6.942	146
	Transparent	147	0.75	0.62	14.678	146
	Non-transparent	147	-0.36	0.73	-6.018	146
Non-autistic	Total	590	0.35	0.46	18.566	590
	Transparent	590	0.89	0.58	37.560	590
	Non-transparent	590	-0.33	0.76	-10.435	590

Note. All values are significantly different from chance at $p > .001$

Table 4.9

Performance against chance on novel deception tasks

Task	*n*	*M* (*SD*)	*Mdn*	*T*	*z*	*r*
WP-3D	590	1.45 (1.08)	2	**115757.00*** **	6.97	0.29
TTL-3D	587	1.88 (1.11)	2	**141164.00*** **	13.47	0.56

Note. WP-3D = Weird Pictures (Development of Deception Detection), TTL-3D = Two Truths and a Lie (Development of Deception Detection). Bold indicates significant value: *** $p < .001$

4.3.4.2. Difference between groups

4.3.4.2.1. Cheating Task

A 2 x 2 mixed Analysis of Variance (ANOVA) was run on the CHR to investigate if groups differed on deception detection on the Cheating Task, with diagnotic group (Autistic / matched non-autistic) as the between subject variable and transparency as the within subject variable. Means and standard deviations of both groups can be found in Table 4.10.

There was a significant main effect of transparency with participants scoring significantly higher in the transparent condition versus the non-transparent condition; F (1,292) = 449.400, $p < 0.001$, $\eta_p^2 = .606$. However, there was no significant effect of group i.e. autistic receivers' deception detection overall was comparable to that of the matched non-autistic sample; F (1,292) = 0.316, $p = .575$, $\eta_p^2 = .001$.

There was a small but significant interaction between transparency and group; F (1,292) = 4.214, $p = 0.041$, $\eta_p^2 = .014$. While both groups performed similarly in the non-transparent condition (t [292] = 1.055, $p = .292$, $d = .123$), non-autistic receivers were significantly better than the autistic receivers in the transparent condition (t [292] = -2.139, $p = .033$, $d = -.250$). Figure 4.10 illustrates the comparison between autistic and non-autistic participants in each transparency condition.

Table 4.10

Descriptive statistics for autistic and (matched) non-autistic participants' CHR

Transparency	Group								
	Across groups			ASC			NT		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Total	294	0.28	0.44	147	0.26	0.45	147	0.30	0.42
Transparent	294	0.82	0.60	147	0.75	0.62	147	0.90	0.57
Non-transparent	294	-0.41	0.73	147	-0.36	0.73	147	-0.45	0.73

Note. ASC = Autistic sample, NT = matched autistic sample *n* = sample size, *M* = mean, *SD* = standard deviation

Figure 4.10

Mean difference in performance on the Cheating Task between groups

Note. Error bars represent one standard error.

4.3.4.2.2. Novel Tasks

Two independent samples Mann-Whitney U Tests were run to see if there was a difference in accuracy on the WP-3D and TTL-3D between the autistic participants and the (matched) non-autistic sample.

The ASC group performed significantly worse than the NT group on the WP-3D, U = 12261.500, z = 2.191, p = .028, r = 0.13. However, the two groups were comparable on their performance on the TTL-3D, U = 11041.000, z = 0.442, p = .658, r = 0.03. Mean, standard deviations and medians for both tasks of each group can be found in Table 4.11.

Table 4.11

Descriptive statistics for novel tasks for both receiver groups

	ASC			Matched NT		
	n	Median	*M* (*SD*)	n	Median	*M* (*SD*)
WP-3D	146	1	1.45 (1.08)	147	2	1.72 (1.03)
TTL-3D	147	2	1.88 (1.11)	146	2	1.95 (1.12)

Note. WP-3D = Weird Pictures (Development of Deception Detection), TTL-3D = Two Truths and a Lie (Development of Deception Detection)

4.3.4.3. Association analyses

A series of Spearman's correlations between deception detection performances of the two groups on the deception tasks and their age, verbal and non-verbal ability, autistic traits, and mentalizing ability were conducted to explore the strength and direction of the associations. Considering that performance was below chance on the non-transparent condition on the Cheating Task, only the transparent CHR is analysed here. Furthermore, as groups were not comparable in their performance in the transparent condition (and the WP), and due to the disparity in their sample sizes, separate associations were run for autistic and non-autistic participants to make sure that trends in the larger non-autistic sample did not influence or obscure any trends in the relatively smaller autistic sample. Correlation coefficients for both groups are presented in table 4.12.

For the ASC group, deception detection ability in the Cheating Task showed a weak but significant correlation with age ($p = .046$, Figure 4.11), however this association was not observed for either of the novel tasks. Unsurprisingly, deception detection ability was highly correlated with verbal ability ($p < .001$) and non-verbal ability ($p = .006$), however only in the Cheating Task. Unexpectedly, autistic traits showed a significant positive correlation with the Cheating Task ($p = .014$) and WP-3D ($p = .003$) that implies that autistic participants who reported higher autistic traits performed better on the deception detection task. Furthermore, deception detection ability correlated only with the AT-Total in the Cheating Task ($p = .008$) and TTL-3D ($p = .012$), but performance on none of the tasks correlated with AT-ToM. Figure 4.12 shows the relationship between mentalizing and the Cheating task.

For the NT group, the correlations revealed different relationships. Deception detection ability in the Cheating Task was highly correlated with age (p < .001, Figure 4.11), but performance on the novel task did not have the same relationship with age. Verbal ability showed a positive significant relationship with the Cheating Task (*p* = .005), and this was true to a lesser extent for the WP-3D (*p* = .033) and TTL-3D (*p* = .012), unlike in ASC. Non-verbal ability on the other hand did not correlate with deception detection ability in any of the tasks in the NT sample, which was a deviation from what was observed for the ASC. Unlike the ASC, autistic traits in the NT showed a negative correlation with deception detection ability, which aligns more with what was expected, although only that in the Cheating Task was significant (*p* = .013). Neither the AT-Total nor the AT-ToM were correlated with deception detection in the Cheating Task (Figure 4.12) or TTL-3D, but accuracy in WP-3D showed a positive relationship with AT-Total (*p* = .001) and AT-ToM (*p* = .011).

Figure 4.11

Relationship between CHR (Transparent condition) on Cheating Task and age

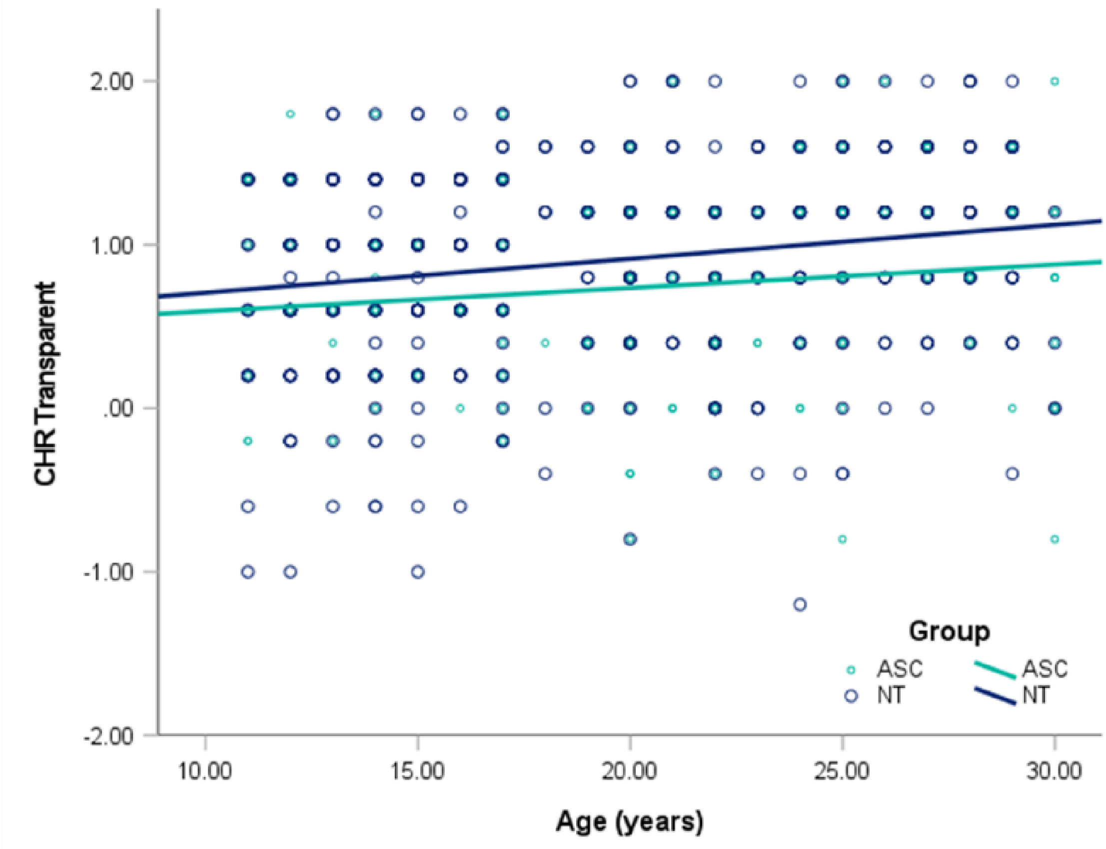
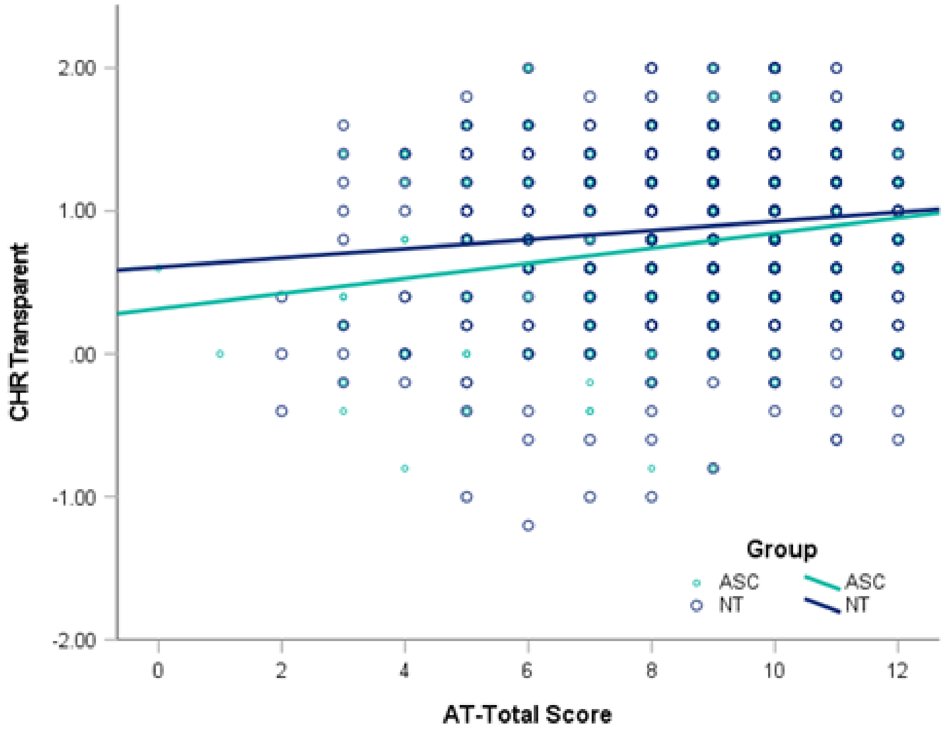


Figure 4.12

Relationship between CHR (Transparent condition) on Cheating Task and mentalizing measures: (a) AT-Total, and (b) At-ToM

(a)



(b)

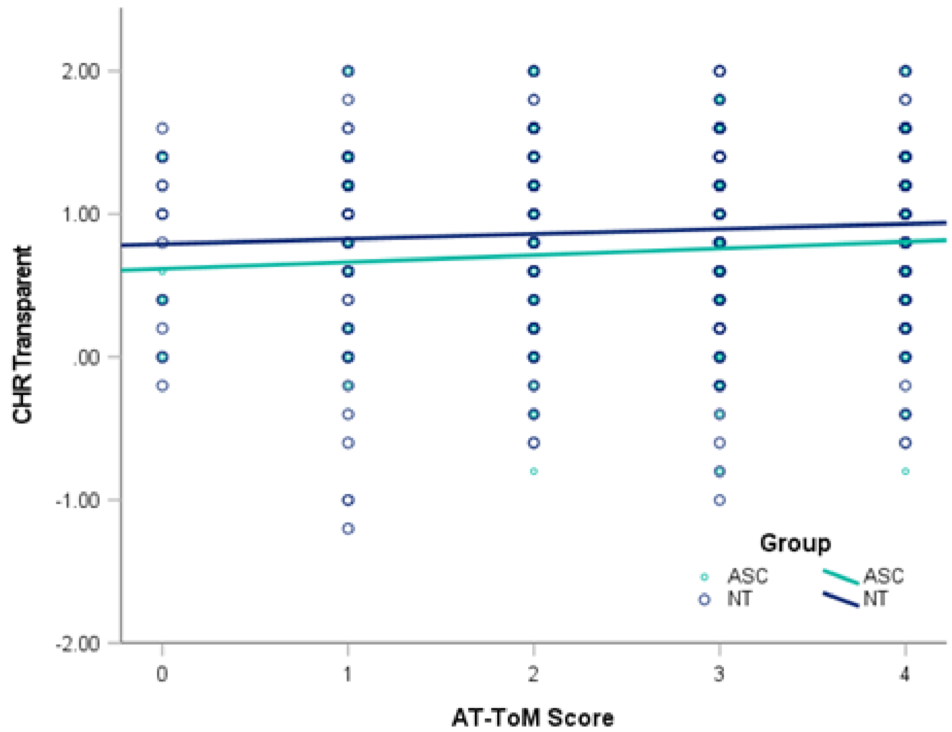


Table 4.12

Correlations between deception detection on the Cheating Task and predictor variables for ASC (above diagonal in grey) and NT (below diagonal)

		1.	2.	3.	4.	5.	6.	7.	8.	9.	*M*	*SD*
1.	CHR-T	—	-.066	.140	.165*	.275***	.227**	.202*	.219**	.115	0.75	0.62
2.	WP-3D	.052	—	.052	.031	.136	.146	.246**	.189	.077	1.5	1.08
3.	TTL-3D	.165***	.022	—	.044	.156	.075	.053	.207*	.067	1.9	1.11
4.	Age	.214***	.070	.063	—	.155	.026	.139	.070	.023	20.94	5.50
5.	STW	.116**	.088*	.103*	.050	—	.466***	.427***	.397***	.245**	77.47	11.49
6.	MaRs-IB	-.015	.067	.036	-.147***	.287***	—	.397***	.475***	.250**	55.60	19.23
7.	AQ	-.102*	-.012	-.026	-.017	-.093*	-.012	—	.307***	.142	30.40	8.94
8.	AT-Total	.074	.131**	.066	.102*	.156***	.296***	-.095*	—	.732***	8.19	2.53
9.	AT-ToM	.037	.104*	.035	.036	.097*	.260***	-.105*	.767***	—	2.80	1.05
	M	0.89	1.6	1.9	18.89	78.69	62.46	20.06	8.87	2.88		
	SD	0.58	1.01	1.05	5.61	9.92	18.51	6.72	2.15	1.07		

Note. Spearman's correlation coefficients for autistic receivers (*n* = 147, 146 for WP-3D) displayed above the diagonal; correlations for non-autistic receivers (*n* = 590, 587 for TTL-3D) displayed below the diagonal.

CHR-T = Corrected Hit Rate for Transparent Condition (Cheating Task), WP-3D = Weird Pictures (Development of Deception Detection), TTL-3D = Two Truths and a Lie (Development of Deception Detection), STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations Total Score, AT-ToM = Frith-Happé animations Mentalizing Score

*$p < .05$, ** $p < .01$, *** $p < .001$ two tailed

4.3.4.4. Predicting deception detection ability

Due to the very different correlations of the two groups, two five-block hierarchical linear regressions were run to predict deception detection – one for each group. Prior to running the analysis, assumptions of multiple regression were tested. Collinearity statistics of VIF and Tolerance were within acceptable limits (Bowerman & O'Connell, 1990; Menard, 2002). An exploration of residual statistics and plots showed that there were no outliers in the data and that the assumptions of normality of residuals, linearity, and homoscedasticity were met.

Only deception detection accuracy on the Cheating Task (transparent condition) was entered as the dependent variable considering it showed the strongest correlations with the predictor variables and because it is a more established measure of deception detection in comparison to the novel tasks, which are being used for the first time. For each group, demographic information – dummy coded gender variables, and English as primary language (L1) – was entered in the first block. To control for any variance due to IQ, proxy measures of verbal (STW) and non-verbal ability (MaRs-IB) were entered in the second block. Next, age was entered in the third block to test if deception detection was predicted by age over and above any variance due to non-verbal and verbal ability. This was followed by autistic traits in the fourth block, and mentalizing ability in the fifth. Since the AT-Total and the AT-ToM were highly correlated, only the AT-Total was entered as the mentalizing variable as it was deemed more representative of the participants' ability to distinguish between mentalizing and non-mentalizing stimuli.

For ASC (Table 4.13), it is evident that Model 1, with just gender and English as primary language, failed to predict any variance in deception detection ability, $F\,(3,143) = .644$, $p = .588$. Model 2 was significant ($F\,[5,141] = 2.897$, $p = .016$) and predicted an additional 8% of variance on the addition of verbal and non-verbal ability, even though verbal ability (β =

.176, $p = .056$) or non-verbal ability ($\beta = .162$, $p = .077$) were only marginally significant standalone predictors. Models 3, 4, and 5 were significant; however, none of them predicted any additional variance in deception detection (Model 3: $F\,[6,140] = 2.639$, $p = .017$; Model 4: $F\,[7,139] = 2.367$, $p = .026$; Model 5: $F\,[8,138] = 2.200$, $p = .031$). Overall, the final model accounted for 11.3% of the variance that was primarily driven by verbal and non-verbal ability.

For NT (Table 4.14), Model 1, consisting of the dummy coded gender variables and English as primary language, did not predict any variance in deception detection ability $F\,(3,586) = .743$, $p = .527$. On addition of verbal and non-verbal ability, Model 2 was significant ($F\,[5,584] = 3.212$, $p = .007$), however it explained only 2.3% of the variance, mainly as verbal ability was a significant unique predictor ($\beta = .162$, $p < .001$). Model 3 was also significant ($F\,[6,583] = 6.511$, $p < .001$), predicting an additional 3.6% of variance with the addition of age ($\beta = .194$, $p < .001$). Model 4 predicted only an additional 0.7% of the variance in deception detection ability, but it was significant ($F\,[7,582] = 6.242$, $p < .001$) with the addition of autistic traits ($\beta = -.085$, $p = .036$) and with verbal ability and age continuing to be significant predictors. Model 5 with inclusion of mentalizing ability ($\beta = .085$, $p = .050$), although significant ($F\,[8,581] = 5.972$, $p < .001$), predict an additional 0.6% variance. Overall, the model accounted for only 7.6% of the variance in deception detection ability, with verbal ability ($\beta = .134$, $p = .002$), age ($\beta = .181$, $p < .001$) remaining unique predictors, and autistic traits ($\beta = -.077$, $p = .059$) and mentalizing ability ($\beta = .085$, $p = .050$) emerging as marginally significant predictors.

Table 4.13

Summary of Hierarchical Regression Analysis predicting deception detection in transparent condition for autistic receivers

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Female	-.27	.25	-.22	-.18	.25	-.14	-.16	.25	-.13	-.12	.25	-.09	-.10	.25	-.08
Male	-.16	.25	-.13	-.03	.24	-.02	-.02	.24	-.01	.03	.25	.03	.05	.25	.04
Eng-L1	.05	.37	.01	-.04	.36	-.01	.01	.36	.00	.00	.36	.00	-.04	.37	-.01
STW				.01	.00	**.18**	.01	.00	**.17**	.01	.01	.15	.01	.01	.13
MaRs-IB				.01	.00	**.16**	.01	.00	**.17**	.00	.00	.15	.00	.00	.12
Age							.01	.01	.10	.01	.01	.10	.01	.01	.09
AQ										.00	.01	.07	.00	.01	.06
AT-Total													.02	.02	.09
R^2	.013			.093			.103			.107			.113		
ΔR^2	.013			.080			.010			.003			.007		
ΔF	.64			6.21			1.61			.48			1.02		

Note. n = 147. Eng-L1 = English as primary language, STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations total score (mentalizing).
Bold indicates marginally significant results at $p < .10$ two tailed.

Table 4.14

Summary of Hierarchical Regression Analysis predicting deception detection in transparent condition for non-autistic receivers

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Female	.09	.14	.08	.12	.14	.11	.14	.14	.12	.10	.14	.09	.10	.14	.09
Male	.14	.14	.12	.15	.14	.13	.14	.14	.12	.10	.14	.09	.11	.14	.09
Eng-L1	-.05	.07	-.03	-.07	.07	-.04	-.04	.07	-.03	-.04	.07	-.03	-.05	.07	-.03
STW				.01	.00	**.16***	.01	.00	**.15***	.01	.00	**.14**	.01	.00	**.13**
MaRs-IB				.00	.00	-.05	.00	.00	-.02	.00	.00	-.02	.00	.00	-.04
Age							.02	.00	**.19***	.02	.00	**.19***	.02	.00	**.18***
AQ										-.01	.00	**-.08***	-.01	.00	**-.08**
AT-Total													.02	.01	**.08**
R^2	.004			.027			.063			.070			.071		
ΔR^2	.004			.023			.036			.007			.002		
ΔF	.74			6.89			22.42			4.40			3.87		

Note. *n* = 590. Eng-L1 = English as primary language, STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations total score (mentalizing).

Bold indicates marginally significant results at $p < .10$, *$p < .05$, ** $p < .01$, *** $p < .001$ two tailed.

4.3.5. Deception detection and peer victimization

Each item on the MPVS-R (Betts et al., 2015) was scored on a scale of 1-3 and all items were summed for each participant to create the total peer-victimization score. Participants' scores could range from 20 to 60, with a higher score indicating more instances of reported peer-victimization.

Independent samples t-tests confirmed that autistic individuals (M = 37.25, SD = 13.21) reported higher peer-victimization than non-autistic individuals in the matched sample (M = 31.76, SD = 9.96); $t(271.45) = 4.024$, $p < .001$, $d = .469$.

To investigate if mentalizing and deception detection predicted peer-victimization, stepwise regressions for each diagnostic group were planned, however, observation of residual statistics and charts showed that the assumption of normality of residuals was violated. Therefore, robust multiple regressions with bias corrected and accelerated bootstrapping (at 2000 samples) were performed instead. For each group, participants' mentalizing scores. i.e., the AT-Total, and deception detection scores, i.e. CHR in transparent condition (CHR-T) were entered into the model as regressors along with background factors such as gender, age, verbal and non-verbal ability. All regression coefficients can be found in table 4.15.

For the ASC group (n = 147), the overall model was significant (F [8,138] = 9.832, $p < .001$) and predicted about 36.3% percent of variance in peer-victimization scores. Surprisingly, the MarRs-IB stood out as the strongest unique predictor of peer-victimization in the ASC group (β = -.316, $p < .001$) with the largest effect size, suggesting that those showing scoring the most on the non-verbal ability measure also reported to being victimized the least. Age was also a unique predictor (β = .191, p = .008) indicating that for the ASC participants, experiences of peer-victimization increased as they grew older. Mentalizing ability, i.e., AT-

Total was not found to be a significant predictor. On the other hand, deception detection, i.e., CHR-T was an extremely significant unique predictor ($\beta = -.251$, $p < .001$), showing that the better the deception detection ability, the lower the peer-victimization reported.

In the NT group ($n = 590$), the overall model was also significant ($F\,[8,581] = 7.490$, $p < .001$) but only predicted 9.3% of variance in reported peer-victimization. Similar to the ASC group, non-verbal ability emerged as significant unique predictor ($\beta = -.146$, $p = .004$). Age was a unique predictor in the NT group as well ($\beta = -.083$, $p = .033$), but the direction of this relationship was opposite to what was observed in the ASC group: reported experience of peer-victimization decreased with age for the NT participants. Unlike the ASC group, AT-Total did predict unique variance in the NT group and had the biggest effect size ($\beta = -.163$, $p = .002$). Deception detection, i.e. CHR-T, however, was only marginally significant as a unique predictor ($\beta = -.082$, $p = .052$), demonstrating that higher deception detection accuracy corresponds to lower reported peer-victimization.

(Rest of page intentionally left blank)

Table 4.15

Summary of robust multiple regression analysis predicting peer-victimization

	ASC				NT			
	B	*SE B*	β	*p*	*B*	*SE B*	β	*p*
Female	5.81	4.26	.218	.145	-2.74	2.73	-.133	.321
Male	6.55	4.23	.249	**.095**	-1.44	2.72	-.070	.586
Age	.46	.16	.191	**.008**	-.15	.07	-.083	**.033**
STW	-.05	.11	-.044	.619	-.01	.05	-.010	.814
MaRs-IB	-.22	.06	-.316	**<.001**	-.08	.03	-.146	**.004**
AQ	-.19	.12	-.126	.132	.09	.06	.061	.105
AT-Total	-.36	.43	-.068	.406	-.78	.23	-.163	**.002**
CHR-T	-5.36	1.44	-.251	**<.001**	-1.45	.74	-.082	**.052**

Note. Unstandardized coefficient and error bias-corrected and accelerated bootstrapped at 2000 samples. Standardised coefficient not bootstrapped.

Female and Male are dummy coded gender variables, non-binary / other is coded as 0 for both variables. STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations total score (mentalizing), CHR-T = Corrected hit rate for transparent condition on Cheating task (deception). Bold indicates significant or marginally significate results.

4.3.6. Deception detection, peer-victimization, and mental health

Mental health issues, operationalised as psychological distress, were measured with the help of the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983), and a total score was calculated by summing all the items (some of which were reverse scored). Participants could have a total HADS score between 0 and 42, a higher score indicating higher levels of anxiety and depression.

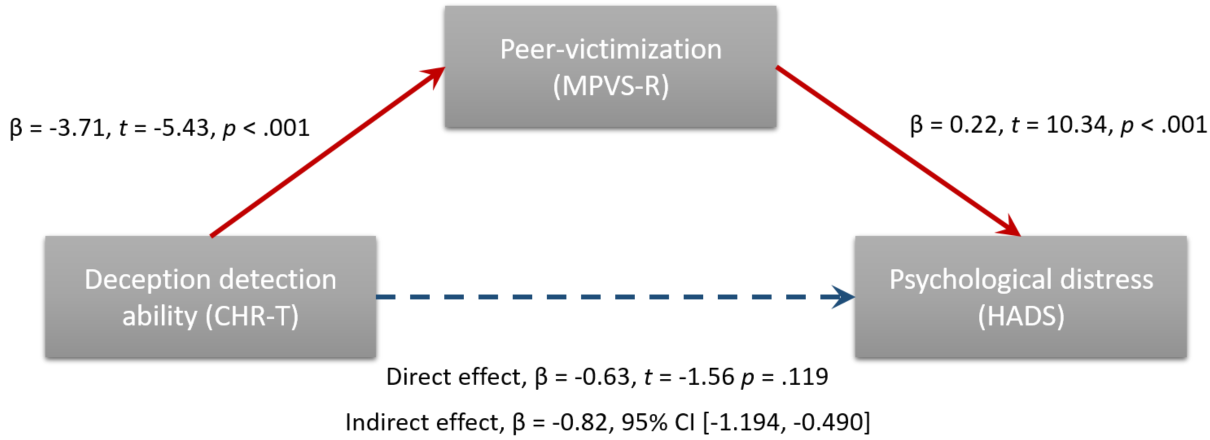
An independent samples t-test confirmed that autistic individuals (M = 18.70, SD = 7.16) reported higher psychological distress than non-autistic individuals (M = 13.97, SD = 6.90) in the matched sample, t (292) = 5.767, $p < .001$, d = .673.

To investigate if deception detection (X) ability indirectly predicted psychological distress (Y), mediated by peer-victimization (M), a mediation analysis using the PROCESS tool (Hayes, 2022) on SPSS was run on the full sample (n = 737). Only the Cheating Task was entered as the measure for deception detection considering none of the novel tasks showed strong unique relationships with the mediator i.e. peer-victimization.

The zero-order effect or total effect of deception detection on psychological distress was significant (β = -1.45, $p < .001$), suggesting that individuals who were worse at detecting deception also reported higher levels of psychological distress. The mediation analysis further revealed an indirect effect of deception detection on psychological distress (β = -.82; LLCI = -1.194; ULCI = -0.490). Considering the confidence interval did not include zero, this effect was considered significant. The direct effect of deception detection on psychological distress however was no longer significant (β = -0.63, p = .119), therefore suggesting that peer-victimization fully and complementarily mediates the negative relationship between deception detection and psychological distress. A summary of the mediation model in represented in Figure 4.13.

Figure 4.13

Mediation model of deception detection as a predictor of psychological distress, mediated by peer-victimization



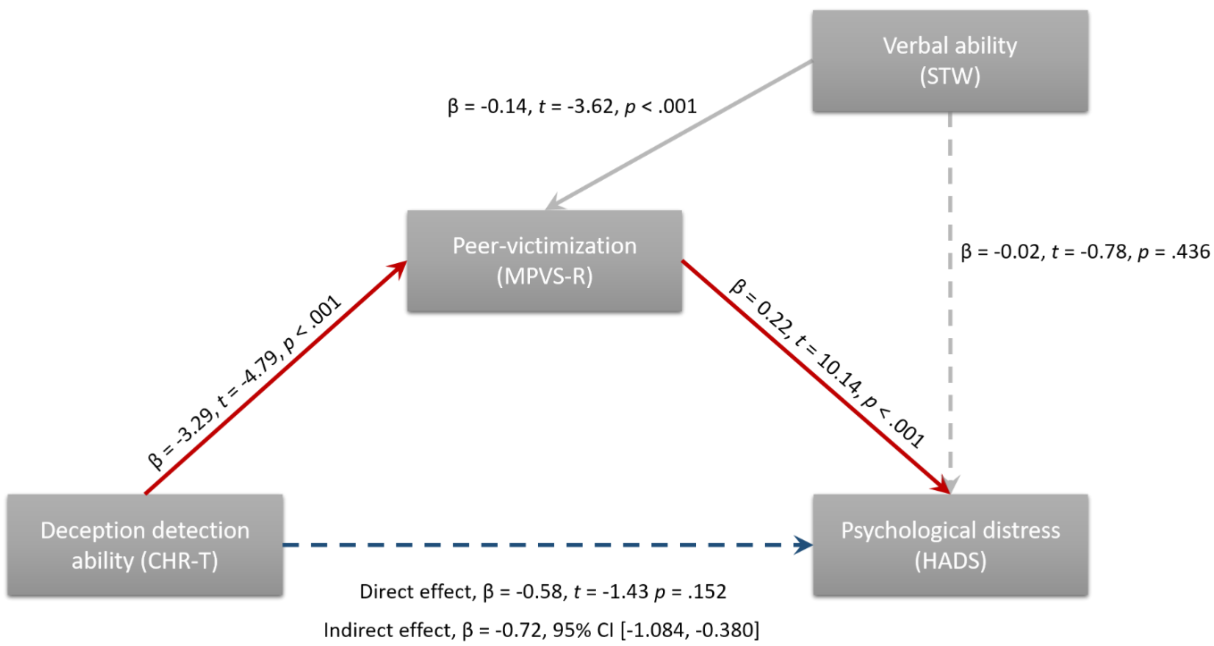
Note. Direct effect between deception detection (X) and psychological distress (Y) represented in blue, indirect effect mediated by peer-victimization (M) represented in red. Solid lines indicate significant relationships, while dashed lines indicate non-significant relationships. The confidence interval for the indirect effect is a bias-corrected and accelerated bootstrapped CI at 5000 samples.

Considering that both the MPVS-R and HADS are self-report questionnaire measures, it was possible that verbal ability may have influenced participants' scores. Thus, the same mediation model as above was re-run while controlling for verbal ability: it was examined if deception detection (X) ability indirectly predicted psychological distress (Y), mediated by peer-victimization (M), but this time verbal ability, i.e. participants' performance on STW, was added in as a covariate (CV).

The total effect of deception detection on psychological stress was significant (β = -1.30, p = .002). Even after controlling for verbal ability, the mediation analysis confirmed that there was an indirect effect of deception detection on psychological distress (β = -.72; LLCI = -1.084; ULCI = -0.380), which can be taken to be significant as the confidence interval does not include zero. The direct effect of deception detection on psychological distress in this case too was non-significant (β = -0.58, *p* = .152), thus further evidencing that there was a full and complementary mediation of the negative relationship between deception detection and psychological distress by peer-victimization. The refined mediation model including verbal ability as a covariate is represented in figure 4.14.

Figure 4.14

Refined mediation model of deception detection as a predictor of psychological distress, mediated by peer-victimization, while controlling for verbal ability.



Note. Direct effect between deception detection (X) and psychological distress (Y) represented in blue, indirect effect mediated by peer-victimization (M) represented in red. Effect of covariant verbal ability (CV) represented in grey. Solid lines indicate significant relationships, while dashed lines indicate non-significant relationships. The confidence interval for the indirect effect is a bias-corrected and accelerated bootstrapped CI at 5000 samples.

4.4. Discussion

In this chapter, I aimed to answer several questions. First, I aimed to investigate if mentalizing developed with age from late childhood / pre-adolescence through adolescence to early adulthood, and if this development trajectory was similar in autistic and non-autistic individuals. As expected, autistic participants performed worse than non-autistic participants on the overall task, although they were comparable on the mentalizing-only subtask (H_1). While performance on the overall task was better with age for the non-autistic participants, this was not true for the autistic participants suggesting that mentalizing ability develops with age in non-autistic participants but not autistic participants. (H_2). The expected interaction between age and diagnosis was not found indicating that even though mentalizing ability appeared to develop differently in the two groups, this difference was not significant (H_3).

Second, unlike what was predicted, mentalizing ability only marginally predicted deception detection ability in non-autistic receivers, and although it significantly correlated to deception detection in the autistic group, it did not predict any unique variance (H_4). Age too predicted variance in deception detection but only in non-autistic receivers, demonstrating that non-autistic individuals get better at detecting deception with age but autistic individuals do not (H_5). Similarly, autistic traits also predicted variance in deception detection ability in non-autistic receivers but not in autistic receivers (H_6), although the autistic group was found to be worse at detecting deception as a group (H_7).

Lastly, I examined the impact of poor deception detection, specifically on vulnerability to peer-victimization and mental health. Autistic participants reported higher peer-victimization (H_8) generally, and while mentalizing ability did not predict peer-victimization, deception detection was a negative predictor of peer-victimization across groups and ages showing that those who are better at detecting deception are less vulnerable to peer-victimization (H_9).

Additionally, deception detection ability negatively predicted mental health issues, and this relationship was fully meditated by peer-victimization (H_{10}), thus suggesting that an increased ability to detect deception that leads to lower peer-victimization could be a potential protective factor against mental health issues or psychological distress.

4.4.1. Mentalizing, age, and autism

A version of the Frith-Happé animations (AT; S. J. White et al., 2011; adapted for online use by Livingston et al., 2021) was used to measure mentalizing ability. The task required participants to attribute mental states to two triangle animations while they moved randomly on screen i.e., random animations (such as drifting), or had some physical interaction i.e., goal directed animations (such as following), or interacted in a way that involved manipulation i.e. ToM animations (such as coaxing).

It was expected that autistic participants would have difficulty in attributing mental states, and therefore would not perform as well on the task as non-autistic participants would. The results revealed that the autistic participants performed worse on the overall task than non-autistic participants, confirming the findings of a plethora of research that establish a marked difference between autistic and non-autistic groups on mentalizing tasks (Baron-Cohen et al., 1985; Schuwerk et al., 2016; S. White et al., 2009; see for Baron-Cohen, 2000a review).

It is however important to note that this difference between groups was primarily as the groups' performance varied in the control conditions and not the ToM condition – specifically the random condition. This goes against past research that have demonstrated group differences in the long verbal description version of Frith-Happé videos (AT-Verbal; Abell et al., 2000; Castelli et al., 2002): non-autistic participants use considerably more mentalizing language to describe the ToM videos than autistic participants. In fact, the results here even contradict the patterns found in other studies that use the objective scoring multiple choice versions of the Frith-Happé animations like the current study where the groups differ primarily in the ToM condition (S. J. White et al., 2011; Livingston et al., 2021). However, the results do align with Moessnang et al's (2020) study, who administered the Frith-Happé animations to the largest autistic sample to date – over 400 autistic individuals – and found that the groups did not differ on the ToM condition.

It is important to acknowledge that this study was administered online and used the version adapted by Livingston et al. (2021) for online use, which deviated from White et al.'s (2011) scoring. More specifically, the online version did not have the MCQ-feelings where participants explicitly have to select from three options what each triangle was feeling after correctly categorizing a ToM video – also the scale on which autistic individuals performed worst in comparison to non-autistic individuals. Compared to Livingston et al. (2021), the non-autistic group in the current study performed similarly – highest accuracy for random animation and similar performance on goal directed and ToM animations. It is the autistic group in the current study that deviates from the trends established in past studies where they are comparable to their non-autistic counterparts on the random and goal directed. Instead, in the current study the autistic individuals perform significantly worse than the non-autistic individuals on the random animations – which drives the group difference.

This group difference in the control condition is not unheard of – Wilson et al. (2021) in a recent meta-analysis found that autistic participants underperformed even in the control condition, and that the effect sizes of difference in the ToM condition were not large (see also Andersen et al., 2022). Wilson et al. (2021) also note that that the group difference in the objective version AT-MCQ as developed by White et al. (2011) is smaller than that in the AT-Verbal where the participants' verbal descriptions are rated as per the criteria by Abell et al. (2000) or Castelli et al. (2002). There is, therefore, a possibility that the AT-MCQ makes it easier for autistic individuals to perform better at the animated task.

This could be due to a few different reasons. First, the objective scoring / categorization in AT-MCQ makes the Frith-Happé animations explicit, much more so than the AT-Verbal. Whereas in the AT-Verbal participants are simply asked to describe what they think is happening in the animation, in the AT-MCQ participants are prompted to think about what sort of relationship the triangles have and if their movements are mentalizing-related. It has been theorized that autistic individuals pass explicit mentalizing tasks using compensatory techniques (U. Frith, 2004) and this could potentially explain why the autistic group perform as well as the non-autistic sample in this study.

The second reason as to why autistic individuals in this sample are better at the AT-MCQ than previous studies could be due to an increased awareness in the sample about what the study is investigating. Online studies of this nature run the risk of self-selection biases (Bethlehem, 2010; Newman et al., 2021), for example it is possible that only those autistic individuals interested in research likely participate and that they also routinely take part in other research studies. This, combined with the awareness amongst autistic participants' about supposed theory of mind difficulties in autism (Gillespie-Lynch et al., 2017) and the widespread use of AT as a tool to investigate mentalizing in autism (Wilson, 2021), makes it probable that the participants are cognisant of exactly what the task is examining. Knowing what the task is about or having attempted it before, may then influence the way they interact with the task (J. Chandler et al., 2015; Hauser et al., 2019). Autistic individuals may compensate – knowing that the task measures mentalizing ability they may attempt to categorize the animations, including the control animations, as ToM more often as opposed to if they were to attempt the task naively. This could lead to them scoring better in the ToM condition due to an elevated hit rate; this reasoning also has the potential to explain the unusually low accuracy in the random condition if they categorize random animations as ToM as well. If this is indeed the reason, the total / overall score on the AT-MCQ will take this into account as better performance on one condition will balance out a worse performance on another.

Therefore, while the results from the ToM condition may not follow trends seen in past studies, the overall scores support previous literature in terms of how mentalizing differs between autistic and non-autistic samples. The developmental trends revealed in this study too partially support previous research. The overall task performance by the non-autistic group showed a strong positive correlation with age, lending support to Klindt et al.'s (2017) large scale online study where they found that the Frith-Happé animations categorization scores (AT-MCQ) improved with age from late childhood to early adulthood in a large non-autistic sample. In fact, the current study improves on their study by including a larger sample of pre-adolescent individuals. However, the findings from the autistic sample here are not in agreement with past literature – Moessnang et al. (2020) found that mentalizing has a positive relationship with age not only in their non-autistic group but also in the autistic group. Then again, Bal et al. (2013) note that age is not as strong a predictor of mentalizing in autistic samples as it is for non-autistic samples. That, along with the fact that the patterns of performance of the autistic group in this study do not completely align with those of previous studies that administered the AT-MCQ in autistic samples, it could explain why this sample of autistic participants do not show observable improvement in mentalizing with age.

Even though the non-autistic group shows a significant relationship with age on the overall task, this trend is not observed for their performance on the ToM condition. Neither Moessnang et al. (2020) nor Klindt et al.'s (2017) report the developmental change in just the ToM condition, making it hard to gauge if these results were unique to the current study, or if this is a concern with AT-MCQ in general where other unexplored or unaccounted for factors, such as compensation, contribute to the performance on the task. In fact, Wilson et al. (2021) comment in their meta-analysis that the Frith-Happé animations do not correlate with other mentalizing tasks and that there is possibility that any individual differences observed on the Frith-Happé animations may not be representative of individual differences in mentalizing ability. Andersen et al. (2022) too conclude that the MCQ may not be a valid way to capture mentalizing ability.

To sum up, the results reveal group differences on the overall AT, suggesting that the autistic participants in our sample did in fact struggle with mentalizing, however it needs to be explored in the future if their higher than usual performance on the ToM condition is due to compensatory behaviour. Similarly, although developmental trends were seen in the non-autistic group and not in the autistic group, the possible concerns with how participants attempted the task mean that any conclusions that can be drawn about the developmental trends observed here (or lack thereof) for either group need to be done so with some caution.

4.4.2. Deception detection and autism

A central aim of this chapter and broader project was to investigate deception detection in autism. Deception detection was measured using a well-used paradigm – the Cheating task (T. R. Levine, 2007–2011) – and the two novel tasks created specifically for this project (Chapter 2). Performance was above chance for all the tasks demonstrating that the tasks were working as they were expected.

The results revealed that autistic and non-autistic participants were comparable on the cheating task overall, but when judging lies from transparent senders – i.e. those who are thought to leak more cues – autistic receivers performed worse than non-autistic receivers. This replicates what Williams et al. (2018) found in an adult sample and extends their findings to a younger sample. However, where Williams et al. (2018) found large effect sizes for the difference between groups, only small to medium effect sizes were found in this study. Given that the sample size in this study was considerably bigger than that of the previous study, this could be an indication that, while autistic individuals do face some difficulties in judging lies, they may not be as weak in deception detection as previously suggested.

Alternatively, the smaller group difference in the transparent condition could be because this study tested younger participants. Deception detection ability was found to develop with age in the non-autistic group, but not so in the autistic group, meaning the gap between the two groups' abilities is likely much more prominent in adults, like the sample tested by Williams et al. (2018). The inclusion of an adolescent sample, where the gap between groups is likely smaller, possibly weakens the overall effect found here. The developmental trends of deception detection are discussed further in section 4.4.4. below.

This diagnostic group difference also emerged in one of the two novel tasks used – the weird pictures (WP), but not the two truths and a lie (TTL). One reason could be that the TTL was an easier deception task – from the group means, participants performed generally better on the TTL than on the WP (although this was not tested directly), and while non-autistic participants were significantly above chance on both tasks, there was a larger effect for the TTL. Additionally, of the two novel tasks, TTL was more contextual – it was more ecologically grounded (vs describing pictures) and receivers were told what the senders would be talking about (e.g., hobbies), and context helps improve deception detection (Blair et al., 2010). Therefore, it is possible that the relative easiness of the task made it less sensitive to the differences in ability between the two groups. Conversely, this could mean that the underlying processes in deception detection in the two novel tasks are different, i.e., different abilities feed into deception detection in the two tasks and autistic individuals struggle with only those affecting performance in the WP. Given these tasks are novel, before concrete conclusions can be drawn from them, more investigation into the cognitive mechanisms feeding into the two tasks is required. A cognitive profile of each of the two task needs to be drawn focussing on more than just mentalizing, which is the sole cognitive ability measured as part of this study, and should include other cognitive skills like emotion perception that has been found to be factor affecting deception detection (Stewart et al., 2019). The validity and effectiveness of novel tasks are discussed further in Chapter 6.

Beyond group differences, autistic traits showed opposite relationships to deception detection in the two groups. Astonishingly, autistic traits showed a significant positive relationship to deception detection in the autistic group, suggesting that the more autistic traits a receiver reported, the better they were at detecting deception in the Cheating Task, which was in direct contrast to what Sowden et al. (2018) and Williams et al (2018) discovered. This unexpected trend was possibly an artefact of verbal ability, however, as both autistic traits and the Cheating Task were highly correlated to verbal ability. This proved to be true as the relationship was no longer significant once verbal ability was controlled for. Autistic traits in the non-autistic group on the other hand did negatively predict their deception detection ability, even after controlling for verbal ability.

This indicates that while autistic individuals may be marginally worse at detecting deception, the degree to which they report having autistic traits does not affect how good or bad they are at detecting lies. However, in the non-autistic group, more autistic symptomatology meant worse deception detection skills. Across the full sample, it likely means that the effect of autistic symptomology on deception detection ability plateaus at a certain point, and likely this point is close to the clinical cut off, and hence the relationship between autistic traits and deception ability becomes non-existent in the clinically diagnosed autistic group.

The most straightforward reason for the difference in deception detection abilities between the two groups is that there is a difference in mentalizing ability, and therefore that the autistic group struggles to read the cues even in senders who are considered transparent. An alternate explanation could be that, as Williams et al. (2018) suggest, deception detection potentially depends on the degree to which one learns social cues by repeated social engagement and communication. Admittedly this is likely to be reduced in individuals with higher autistic traits due to less social motivation (Dubey et al., 2015), and increased social anxiety (Bejerot et al., 2014; Kuusikko et al., 2008). It will be interesting to see in future research if the amount of social experience and opportunities to learn social cues that autistic individuals have mediates the relationship between autistic traits and deception detection.

4.4.3. Deception detection and mentalizing

Logically, mentalizing must play a role in deception detection – in order to be able to judge when one is being deceived, one must be able to gauge the other's mental states such as intentions, beliefs, and knowledge. Empirically, there have been only a couple of studies actually testing this, and one of the aims of the current study was to investigate this relationship.

Results revealed that even though mentalizing strongly correlated with deception detection ability in the autistic group, it did not predict any unique variance once verbal ability was controlled for. In the non-autistic group, mentalizing ability only marginally predicted deception detection ability, in contrast to Stewart et al.'s (2019) study, where they found a strong link between mentalizing and deception detection. The positive but non-significant correlation found for the non-autistic group instead support the outcomes of Williams et al. (2018), who, like the present study, also used the cheating task (T. R. Levine, 2007–2011) to measure deception detection ability and the Frith-Happé animations to measure mentalizing ability and found no correlation between the two. However, they only investigated this relationship in a non-autistic sample. The current study, in addition to testing a sample much larger than both Stewart et al.'s (2019) and William et al.'s (2018), also expands on William et al.’s (2018) work by examining the relationship between mentalizing and deception detection in autistic individuals as well. To the best of my knowledge, this is the first study to do so, especially in a developmental population.

One reason for the weak relationship between mentalizing and deception detection found here could be that deception, like mentalizing, is characterized by dual processes – explicit and implicit (Apperly & Butterfill, 2009). It can be intuited that deception production would require a more active, intentional, and resource demanding explicit mentalizing process, while deception detection has to be more intuitive and implicit as, outside of laboratory scenarios, people are not forewarned that they are going to be deceived. For that reason, the AT-MCQ, which is a fairly explicit mentalizing task, may not predict deception detection ability efficiently. The more subjective AT-Verbal might be able to detect trends better considering it is comparatively more implicit.

The second reason is related to the concerns regarding the differing trends in the Frith-Happé animations from previous studies. As mentioned earlier, there are concerns that participants’ performance on the AT-MCQ could have been influenced by a self-selection bias and non-naivety of the sample and the measure of participants’ mentalizing ability could have possibly been confounded by compensation. In that case, the weak relationship between the mentalizing and deception detection ability in this study cannot be taken as a confirmation that such a relationship does not exist in reality. To draw conclusive results, future studies must control for other confounding variables such as compensatory mechanisms when using the AT-MCQ, or instead use implicit measures to capture mentalizing ability to confirm if deception detection is associated with the ability to infer mental states.

4.4.4. Deception detection and age

Given that mentalizing was thought to be an important cognitive function driving deception detection, it was expected that deception detection will develop with age – the older one was, the better they would be at detecting lies. Results revealed that for the non-autistic group, age was the strongest predictor of deception detection amongst all the factors tested, showing that deception detection ability did indeed develop with age. This supports the sparse research in this area – both DePaulo et al. (1982) and Shaw and Lyons (2017) found that older participants were better at detecting deceit than younger participants. Though the findings from these studies could not be taken as conclusive evidence due to possible confounding factors (see introduction – section 4.1.3), the current study used a well-established deception paradigm to map the developmental changes in deception detection from late childhood, through adolescence, to early adulthood.

While the non-autistic group shows a definite effect of age on deception detection abilities, the autistic group shows no such developmental trend. These findings can be interpreted as to mean that where non-autistic children grow to become better lie detectors, autistic children likely do not. Given that autistic individuals were generally found to be relatively worse at deception detection, this potentially leaves them vulnerable to being deceived, more so than their non-autistic counterparts are, in adulthood.

One possible reasoning for these findings borrows from child studies of mentalizing and deception (production) – that deception is an everyday marker of mentalizing (M. Chandler et al., 1989; Hala et al., 1991; Sodian, 1991). The developmental trends in deception detection in both groups mimic the predictions and results for mentalizing – autistic individuals are relatively weak at it, and it develops with age, at least in non-autistic individuals (e.g.: Devine & Hughes, 2013). For autistic individuals it aligns with literature that suggests that while they are able to pass tests that require explicit mentalizing as they grow older, they continue to struggle with tests that are more reflective of real-life utilization of mentalizing (e.g. Moran et al., 2011), likely due to difficulties in implicit mentalizing (Schneider et al., 2013). This makes it even more probable that the reason that a stronger relationship between deception detection and mentalizing was not observed was due to the probability that the Frith-Happé animations were not tapping into a mentalizing process that was implicit enough to affect deception detection.

4.4.5. Other factors affecting deception detection

The associations drawn for both groups revealed some interesting trends for deception detection. Deception detection ability in autistic and non-autistic individuals appeared to have very different drivers. The only common factor that seemed to predict even a fraction of variance in both groups was verbal ability. This is understandable given that the deception tasks required well-developed verbal comprehension, considering that the tasks only tested verbal lying and not non-verbal deception (such as when a player exaggerates the extent of their injury during a game to earn a penalty shot). What was surprising was that while the model for autistic participants was significant, none of the individual predictors was uniquely able to predict variance in deception detection. One reason why this generally happens is multicollinearity, i.e. extremely high correlation between predictors. Since this was not the case here, the only explanation is that a combination of factors – likely verbal and non-verbal ability – were together responsible for predicting most of the variance by the model for autistic individuals. However, it needs to be noted that the overall models for both groups predicted only a small amount of variance in deception detection. This shows that there could be other factors beyond those that were examined here that drive deception detection.

One factor that has been observed to affect deception detection is thought to be sender demeanour, i.e. how honest or trustworthy a sender appears regardless of their actual veracity (T. R. Levine, 2016; T. R. Levine et al., 2011; Semrad et al., 2019). However, in Chapter 3 it was seen that controlling for demeanour between veracity types possibly negates this effect of demeanour on deception production success. Furthermore, because demeanour is a property of the individual sender, and deception detection ability is a property of the receivers (T. R. Levine, 2010), it is unlikely that demeanour would predict individual differences in deception detection abilities of the receivers.

Another factor effecting deception detection, especially when studying differences between groups, could be intergroup bias. As briefly mentioned in Chapter 1, group identification affects credibility (Rozmann & Nahari, 2021) and perception of truthfulness of senders (Fan et al., 2022). There is also evidence that autistic individuals are better at communicating with other autistic individuals (Crompton, Ropar, et al., 2020; G. L. Williams et al., 2021) and that autistic individuals are better at attributing mental states to those who are also autistic versus those who are not (Edey et al., 2016; Partington et al., in prep.). In fact, the Double Empathy Problem (DEP) suggests that the social-communication difficulties that autistic individuals face are not solely due to autistic cognition but because autistic and non-autistic people have very different experiences of the world and hence there is breakdown in communication between the two groups (D. E. M. Milton, 2012). It is therefore possible that when testing for deception detection in autistic individuals, they are disadvantaged because the senders are all non-autistic and they struggle to read the cues to deception even though these senders are transparent. Therefore, to see if deception detection may be affected by intergroup bias or the DEP, it is essential to test autistic receivers' ability to detect deception not only from non-autistic senders but from autistic senders as well, which I will do in Chapter 5.

4.4.6. Impact of deception detection on peer-victimization and mental health

A final aim of this study was to investigate the impact of poor mentalizing, and specifically deception detection, on quality of life, and more specifically if it affected how vulnerable one was to peer-victimization, and if a reduced ability to detect deceit could indirectly influence mental health and psychological distress.

Mentalizing ability predicted vulnerability to peer-victimization in the present study but only for the non-autistic group. This supports Shakoor et al. (2012), who found that poor performance on a battery of mentalizing tasks predicted the possibility of being a victim to bullying, and Wahyuningsiha and Novitasari’s (2016) claims that theory of mind negatively correlated with peer-victimization in Indonesian adolescents. Even though it is well accepted that autistic individuals struggle with mentalizing and peer-victimization is much higher in autistic individuals than in their non-autistic peers, this study is one of only studies that directly investigates the effect of mentalizing ability on vulnerability to bullying in autistic individuals. It was found that for the autistic individuals, mentalizing ability was not related to self-reported peer-victimization. Although this is in contrast to van Roekel et al. (2010) who found that autistic adolescents mentalizing ability negatively correlated to their victimization reported by their teachers, it does align with the lack of correlation they found between the mentalizing and self-reported victimization.

While mentalizing did not predict peer-victimization, the ability to detect deception was found to be an extremely strong predictor of peer-victimization in the autistic group, but only a marginally significant predictor in the non-autistic group. Past studies have indicated that deception *production* abilities or prevalence are linked to bullying *perpetrator* behaviour (Keating & Heltman, 1994; Maboyana & Sekaja, 2015), and vulnerability to deception has been identified as a theme in qualitative interviews with parents of autistic children

(Sofronoff et al., 2011). However, this is the first study, to the best of my knowledge that definitively establishes the link between deception *detection* ability and bullying *victimization*.

While the non-autistic group shows a strong significant relationship between mentalizing and victimization, and a marginally significant relationship between deception detection and victimization, the opposite is true for autistic individuals – they show a strong connection between deception detection and victimization but none so between mentalizing and bullying. This can be explained by the fact that mentalizing was found to be a predictor of deception detection ability in the non-autistic group, unlike the autistic group. Therefore, it is possible that mentalizing and deception detection likely explain some overlapping variance in victimization in the non-autistic group whereas in autistic individuals, deception detection abilities that are unaffected by mentalizing predict victimization.

Both the autistic and non-autistic groups showed a change in reported victimization with age, however in different ways: in non-autistic individuals, bullying victimization decreased as the participant got older, whereas the autistic group reported more peer-victimization with age. This could be because deception detection skills, one of the predictors of victimization, do not improve with age in autistic individuals as they do in non-autistic individuals (as evident in section 4.4.4). So, it is possible that while non-autistic individuals grow up to be better detectors and therefore are less vulnerable to victimization than when they are younger, this does not happen for autistic individuals.

However, while this explains why victimization does not decrease in the autistic group like the non-autistic group, this does not explain the *increase* in reported peer-victimization in autistic adults. Correct perception of bullying behaviours has been linked to understanding of social behaviour and theory of mind in autistic individuals (van Roekel et al., 2010) and theory of mind is thought to develop with age even in autistic individuals (Moessnang et al., 2020). Taken together, this could potentially mean that older autistic individuals have a better understanding of when they are being bullied and therefore they report being more vulnerable than younger autistic individuals. Alternatively, the reason for the increased peer-victimization in adulthood for the autistic group could be that the protective factors present during adolescence, are no longer present in adulthood, or are present in reduced capacities. For example, autistic adolescents and young adults have similar difficulties in making friends (Eaves & Ho, 2008; Locke et al., 2010), but adolescents befriend adults at schools, such as teachers, who they often confide in when faced with bullying (Humphrey & Symes, 2010). Autistic adults, on the other hand, often feel isolated and overwhelmed with the independence that comes with being adults (Simmeborn Fleischer, 2012). Therefore, the weaker inbuilt support structures in adulthood, such as parents and teachers who might have protected against prolonged victimization in late childhood and adolescence, could lead to higher reports of peer-victimization in autistic adults.

While deception detection does seem to affect peer-victimization, an interesting revelation was that for both groups, non-verbal ability appeared to be a very strong contributor to predicting variance in the reported peer-victimization as well. Matrix reasoning tasks, such as the one used in thus study to measure non-verbal ability (Chierchia et al., 2019), primarily capture fluid intelligence (Sattler, 1988) that contributes to reasoning and deduction, such as being able to form analogies (Bitsika & Sharpley, 2018). Fluid intelligence has been thought to affect psycho-social adaptation (Huepe et al., 2011) and has shown direct connection to victimization (Liu et al., 2014), such that low fluid intelligence predicts higher propensity to being victimized. Hence, it is not surprising that non-verbal ability in this study predicts peer-victimization. However, it must be emphasised that deception detection predicts unique variance in reported peer-victimization, marginally so for non-autistic individuals, even after the variance due to fluid intelligence has been accounted for, showing that the effect of poor deception detection on greater vulnerability goes beyond any psychosocial maladaptation explained by fluid intelligence.

Ranick et al. (2013) alluded to the fact that deception detection affects peer-victimization, which further impacts mental health, especially in autistic children. However, this is the first time that this relationship has been empirically investigated. Deception detection showed a strong negative relationship with psychological distress, i.e. those who were better at detecting deception reported less psychological distress. This relationship was also fully mediated by peer-victimization. In other words, the psychological distress caused due to being poor at deception detection was actually due to the increased vulnerability to peer-victimization caused by the inability to detect lies, which in turn causes an increase in psychological distress.

The results indicate that that peer-victimization affects psychological distress and mental health issues like anxiety and depression – this is supported by the plethora of past research (Armitage, 2021; Cappadocia et al., 2012; Chang et al., 2013; Shtayermman, 2007; Verkuil et al., 2015). However, this is the first study to map the relationship between deception detection, peer-victimization, and mental health together. While the findings show an extremely strong relationship, this exploration was based on very little past empirical evidence as the relationship between deception detection and bullying victimization in of itself was tested for the first time in the current study. Therefore, it is vital to replicate these results before drawing strong conclusions. Even though the findings hold true despite controlling for verbal ability, it would be beneficial to investigate these relationships using measures that are not self-report and therefore are not affected by participants' verbal ability, such as parent or teacher reports in the case of the younger participants for peer-victimization. Reports from clinicians / counsellors / mental health professionals for psychological distress or mental health difficulties for adolescents and adults alike similarly may also strengthen the evidence found here further, although there does not appear to be any significant relationship between verbal ability and responses on the self-report mental health questionnaire in this particular group of participants.

Furthermore, while the self-report measure used to record peer-victimization used here has been implemented in adult samples before (see: S. Joseph & Stockton, 2018), the MPVS-R (Betts et al., 2015) was created with children and adolescents in mind and as such does not include items that specifically capture workplace bullying that adults are more likely to face. Therefore, future studies could also attempt to capture the nuances of adult-specific victimization and map its relationship with deception detection and psychological distress.

4.4.7. Implications

The developmental trajectory of deception detection ability demonstrates that while autistic individuals may learn to compensate on mentalizing tasks as they grow older, they face difficulties in tasks more akin to the real-world where they have to make judgements about people's intentions, such as in deception detection tasks. While the mentalizing measure used in this study only demonstrates weak developmental trends and group differences, one possible reason as discussed earlier is the overuse of the task in the autistic population. Development of newer, more novel tasks that require more flexible and real-world implementation of mentalizing, are the need of the hour. Considering that deception (production) tasks have been used as tasks of ToM development in children, deception (detection) tasks such as the ones used here could be adapted to study development of mentalizing in a novel and more everyday context. The findings from this study provide an optimal opportunity – the trends of deception detection and group differences mimic what would be expected from a valid mentalizing task. While theoretically sound, validity would need to be established by testing non-autistic naïve samples on both this task and a lesser-known mentalizing task.

Beyond research relevant implications, this study also highlights the consequences of having weak deception detection abilities. Autistic individuals have reported engaging in fewer social interactions, and making fewer friends, along with being the target of more bullying behaviour in comparison to their non-autistic peers (Wainscot et al., 2008). In fact, having fewer friends has been found to be linked to how much autistic individuals are bullied (Cappadocia et al., 2012). This becomes especially problematic because, if an autistic individual is unable to gauge when they are being lied to, there is a lesser chance of a friend standing up for them or translating the intentions of the prospective bully, leaving the autistic individual even more open to abuse, and therefore psychological distress and mental health issues. Therefore, there is a need to explore ways to improve deception detection abilities in autistic individuals, which may alleviate some of their psychological distress that is caused by deception related peer-victimization.

Whether or not deception detection training works is controversial – some reviews suggest that such training only improves detection ability because participants are hyperaware of being lied to (T. R. Levine et al., 2005); however a more recent meta-analysis has shown promising effectiveness of deception detection training (Driskell, 2012). More interestingly, Ranick et al. (2013) has shown that autistic children can be trained to detect deception and that the ability generalises across new interactions and types of lies. Since the premise of developing the training program was to help autistic individuals avoid situations where they can be targets of bullying in which the perpetrators deceive them, this study helps put more weight behind the need to further develop, implement, and adopt deception detection training into mainstream intervention. This study also underlines the need for such training and interventions to be extended to adults considering that autistic adults are much more vulnerable in terms of being bullied due to not being able to grow into better deception detectors. That said, there are some concerns regarding interventions that promote further compensation in autistic individuals, which in turn cause further mental issue, and these will be discussed in Chapter 6.

In addition to implications for the autistic community, this study shows that deception detection is an ability that develops with age and as such, adolescents are less equipped to detect lies and are more vulnerable to peer-victimization. Bullying and peer-victimization is at its peak during adolescence (Nansel et al., 2001). Considering the mental health implications of victimization in adolescence (Chang et al., 2013; Coggan et al., 2003; Ford et al., 2017; Le et al., 2017), this study provides an evidence-based argument to support the introduction of deception detection training to anti-bullying interventions. However, such programmes need to be developed with caution as people good at one deceptions skill are often good at the other, i.e., those who are good at detecting lies are also good at producing lies (Wright et al., 2012). Therefore, it may be prudent to identify those who are most vulnerable to peer-victimization and offer them such training separately to avoid teaching bullying perpetrator to lie better.

4.4.8. Limitations

Apart from the methodological concerns that have been discussed throughout the discussion, some overarching limitations need to be considered when interpreting the findings of this study.

The study took place online, and as with all online studies, there was concern regarding data quality, fraudulent participants, and technical issues. These issues were dealt with using a strict eligibility and exclusion criteria and this led to extensive attrition rates. It has been seen that if precautions are taken, these issues can be mitigated (Gagné & Franzen, 2021) and that results from online and offline tasks often show similar results (Prissé & Jorrat, 2022; Weigold et al., 2013).

An issue with recruiting a clinical population online was that there was no easy way to check that participants met the diagnostic criteria. This was compounded further by the use of self-report questionnaires for autistic traits instead of other diagnostic tools that are based in observation and interview. Self-report questionnaires generally suffer from criticism as they tend to be affected by notions of impression management and open the possibility for exaggerated responses (Paulhus & Vazire, 2007). However, it has been found that self-report measures for variables like peer-victimization are better than reports by others (e.g. parent , teacher) as there is a tendency to miss or overlook instances of victimization (Holt et al., 2008); this is true for autistic individuals as well (R. E. Adams et al., 2014).

The need for online testing arose due to the pandemic that made it impossible to run a more traditional lab-based study. Although this allowed for the large sample size in this study, the pandemic could have potentially influenced our data in other ways as well. The study took place at a time when people's social and school/work life had been severely affected over the previous two years – Wave 1 data was collected summer of 2021 and Wave 2 and adolescent data was collected spring of 2022; the last pandemic related restrictions were lifted in July 2021. While this might, optimistically, not have introduced much variability in their performance on the cognitive and behavioural tasks, the self-report measures may have been biased by the atypical experiences of the pandemic. For example, low social contact could have meant that the amount of victimization reported was lower than at a time prior to the pandemic. I attempted to mitigate this by increasing the standard instruction to report peer-victimization from the past year to two years. Similarly, mental health and psychological distress could have been affected by individual experiences of the pandemic and lockdown related isolation – some people might have experienced an improvement, such as due to reduced opportunities to be victimized, while for others the pandemic meant increased psychological distress. However, it is hoped that the large sample helps overcome the noise introduced by these factors.

As a final point, this study investigated developmental trends in a cross-sectional sample, and therefore the causal relationship between the development of deception detection abilities and bullying victimization need to be interpreted keeping in mind individual differences in experiences and abilities. However, a cross-sectional design helped collect data from a large sample and made it possible to investigate developmental trends across a wide age range within a relatively short span of time, helping us control for atypical environmental factors such as the experience of being in a global pandemic. Nonetheless, it will be interesting to see if the trends observed here hold in a longitudinal design as well.

4.4.9. Supplementary findings based on a sample defined by AQ cut-offs

As mentioned in the previous section, one of the drawbacks of online data collection was that there was no way concrete way to crosscheck the participants' diagnosis. While there were multiple safeguards to prevent wrongful grouping of participants, an alternative way to be more confident in our groupings was to make sure that the participants meet the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, et al., 2001) cut-offs that are typical for their groups. Therefore, all the hypotheses were retested with AQ-defined groups.

Woodbury-Smith et al. (2005) found that, for a clinically diagnosed / referred sample, the highest sensitivity and correct classification was at the threshold value of 26, and advise using this as a cut-off to limit false negatives instead of the original 32 suggested by Baron-Cohen et al. (2001). However, they recommend continuing using the original AQ cut-off of 32 to screen the general population to minimise false positives. Therefore, only those autistic participants who scored 26 or above and only those non-autistic participants who scored below 32 were included. The results for the additional analyses can be found in Appendix A.

Exploring the way this new AQ-defined sample is different from the original sample, it is evident that the autistic group is greatly reduced in number – approximately a third of the original autistic group is excluded. The mean AQ scores for the autistic group is understandably higher in the AQ-defined sample than the original sample, while that of the non-autistic group has gone down. Of note is the fact that the AQ-defined autistic group also has higher verbal and non-verbal ability scores, demonstrating that excluding the autistic participants with the lowest AQ scores has also potentially excluded those from the autistic group with the lowest verbal and non-verbal ability.

The predictive models for deception detection and peer-victimization in both the original sample and AQ-defined sample return strikingly similar results and this is true for both groups. The only difference that emerged was that while predicting peer-victimization, unlike in the original sample, autistic traits surprisingly came forth as a significant unique predictor in the model for AQ-defined non-autistic group. However, this could be a by-product of using the AQ to quantify autistic traits: socially anxious individuals have been found to score highly on the AQ (Tonge et al., 2016). The link between social anxiety and peer-victimization has been well documented (e.g. Pontillo et al., 2019), therefore these results could be reflecting a mutual relationship of the AQ and peer-victimization to social anxiety. Nevertheless, as social anxiety was not measured or controlled for in this study, this needs to be investigated directly before any conclusions are drawn.

The main departure from the original findings in the AQ-defined sample is that the direct group comparisons between autistic and non-autistic group were no longer significant. For mentalizing, the autistic group's surprisingly weak performance in the random control condition, which seemed to drive the marginally significant group difference in the original sample in the first place, were comparable to the non-autistic group, therefore potentially leading to the overall group differences on the mentalizing measure to become non-significant in the AQ-defined sample. This is likely due to a combination of the autistic group performing better in the AQ-defined sample than in the original sample and the matched non-autistic group having a worse performance than that in the original sample. These results therefore show that autistic individuals in this sample are comparable to non-autistic individuals in mentalizing. While unexpected, this is not unheard of. In fact, some past studies have failed to find group difference in the performance on the Frith-Happé animations (Moessnang et al., 2020; Veddum et al., 2019; R. L. Young & Brewer, 2020). That said a large majority of studies using the animations do show a significant difference between groups, as found by Wilson et al. (2021) in their meta-analysis of 33 studies.

Group comparisons for deception detection generated similar results: although the AQ-defined autistic group did perform slightly worse than the corresponding non-autistic group on the Cheating task, this difference was not significant. This is in contrast to the original sample where the autistic group in the original sample was significantly weaker than the non-autistic group. Deception detection, as the premise of this research suggests, has hardly been studied and the findings from the AQ-defined sample contradict that of the only study that uses the same stimuli in autistic adults (D. M. Williams et al., 2018). That said, a recent study by Brewer et al. (2022) found that autistic people were no less capable of detecting dodgy behaviour than the non-autistic participants, however they did find that dodgy-ness was affected by individuals mentalizing ability. Following from that, considering that this AQ-defined sample did not yield differences in mentalizing, it is understandable that the difference in deception detection observed in the original analyses was no longer present in this AQ-defined sample.

In another difference between the original findings and the AQ-defined findings, the autistic people in the AQ-defined sample reported similar levels of peer-victimization as non-autistic individuals. This could be due to both the autistic group reporting more victimization in this cut down sample or the non-autistic group reporting more instances of victimization. However, this goes against a plethora of research demonstrating that rates of peer-victimization faced by autistic individuals is radically higher than their non-autistic peers (Libster et al., 2022; Rowley et al., 2012; Schroeder et al., 2014; Symes & Humphrey, 2010; Ung et al., 2016), including a population based investigation (Øksendal et al., 2019).

One reason that these significant group differences found in the original sample and widely reported in the literature (in the case of deception detection and peer-victimization) are lost in the AQ-defined sample could be a general loss of power. The autistic group was reduced by a third, making it more likely than an effect is not detected in this sample and that the analyses on this group is less powerful than those on the original sample. Considering that the AQ is not a diagnostic tool but a psychometric test to record levels of autistic traits, a possible way to preserve power while also maintaining confidence in the sample's genuineness would be to use the AQ as a continuous measure across the sample in any future studies where there are doubts regarding the autism diagnosis of participants. Albeit, this will answer a slightly different question, one about the variables relationship to autistic traits rather than autism itself, it still provides an alternative to the use of AQ to determine autism diagnosis while still preserving power, enabling one to detect effects that would otherwise be too small to detect with smaller groups.

Furthermore, it must be acknowledged that excluding autistic participants who score lower on the AQ could have led to eliminating autistic individuals who may still be autistic with potentially high autistic traits. There is evidence of high false negatives with the AQ – Ashwood et al. (2016) found that more than 64% individuals in an autism clinic who scored below the AQ cut-off were actually diagnosed with autism. And while the lower cut-off of 26 used here, suggested by Woodbury et al. (2005), boasts of a lower rate of false-negatives than the original cut-off of 32 by Baron-Cohen et al. (2001), it still only has the potential to correctly classify only 83% of diagnosed autistic individuals.

Additionally, lower AQ scores by a certain proportion of the excluded autistic sample may not have been singularly due to presence of fewer autistic traits or less “severe” autistic symptomatology. It could also potentially be due to reduced levels of insight: Past research has suggested that autistic individuals may not have sufficient insight about their difficulties and may have lower self-awareness (U. Frith & Happé, 1999). The AQ, being a self-report measure, would require adequate self-awareness about one’s own difficulties. Further, there is evidence that autistic individuals with higher IQ are more aware of their own social difficulties and rate themselves lower on competencies (Capps et al., 1995; Huang et al., 2017; Vickerstaff et al., 2007) than those with lower IQ. This would indicate that individuals with higher IQ will score higher on the AQ due to their awareness of their own social difficulties, while those with lower IQ would score lower on the AQ. Therefore, the fact that the AQ-defined sample have higher verbal and non-verbal ability than the original sample may suggest that those autistic participants with relatively lower IQ, and potentially lower self-awareness of their own difficulties, were excluded in the process. This further adds to the argument that ‘cleaning’ the data on the basis of the AQ might have its own set disadvantages that need to be weighed against the advantages it provides.

However, still an alternative explanation is that the general increase in performance in autism group in the AQ-defined sample is due to the removal of potentially noisy data such as from participants who did not actually pay attention during the study, did not provide authentic data, or were simply not invested in the study. This is a likely hazard of online testing, despite multiple safeguards. It is also possible then that there are in fact no differences between autistic and non-autistic individuals in mentalizing and deception detection abilities, and in how much victimization they face. Nevertheless, considering the large body of literature suggesting otherwise, the presence of such differences in the original sample, and the loss of power in the AQ-defined sample, a recommendation would be to replicate results found in this study within samples where the diagnosis of the participants can be confirmed and cross checked with more dependable measures such as the Autism Diagnostic Observation Schedule (Gotham et al., 2007).

4.5. Conclusions

It is apparent from this study that deception detection, under-researched in autism as it may be, could play an important role in understanding and improving quality of life for autistic individuals. Autistic individuals can detect deception; however, there is a possibility that this ability may not be comparable with their non-autistic counterparts. Furthermore, this ability does not seem to improve with age in autistic individuals, as it seems to do for non-autistic individuals. Even so, this study demonstrates the necessity of paying more attention to deception detection skills as it has an impact on vulnerability to peer-victimization, especially in autistic individuals, and through it, on mental health.

Further investigation using implicit tasks for mentalizing, and age-appropriate measures to record peer-victimization for adults would help draw stronger conclusions regarding the relationships found. This study also underlines some issues with over-usage of popular tasks, and highlights the opportunity to develop newer and more innovative tasks to test mentalizing in everyday scenarios – such as using deception detection as a measure of mentalizing.

Chapter 5. Deception and the Double Empathy Problem

5.1. Introduction

Consider this: you are visiting a country where the culture and language are unknown to you. Your lived experience is very different from the people of this country and there are nuances in social interaction that you miss because you do not understand the language or the culture. How likely is it that you will be able to successfully navigate a social situation with a stranger in these circumstances, let alone gauge if this person is being genuine and not lying to you? Your answer will probably be ‘highly unlikely’.

An idea that has gathered steam in the autism community is that being autistic is similar to the situation above – being in a world where the majority people have a very different experience and therefore are hard to communicate with. Damien Milton, an autistic autism researcher says: “when people with very different experiences of the world interact with one another, they will struggle to empathise with each other” (D. Milton, 2018). He calls this the Double Empathy Problem (DEP), and suggests that the difficulties autistic individuals face in social navigation are not only because of autistic cognition, but because of a collapse of mutual understanding, empathy and reciprocity between autistic and non-autistic individuals, who have inherently different experiences (D. E. M. Milton, 2012). So, is it possible that any difficulties in detecting deception faced by autistic individuals could be because they are being asked to judge deception from non-autistic senders? Is it possible that if the senders too were autistic, autistic individuals would be as good at detecting deception as non-autistic individuals are from non-autistic senders?

In this chapter, I will explore if the Double Empathy Problem is a potential reason why autistic individuals show difficulty in deception detection. I will investigate whether identifying as autistic makes it harder for receivers to detect lies from non-autistic senders, and whether this is true in the opposite direction as well: whether non-autistic receivers also show similar difficulties in detecting lies from autistic senders.

5.1.1. What is the Double Empathy Problem (DEP)?

The Double Empathy Problem (DEP; D. E. M. Milton, 2012) challenges the widely held belief that the social and communication difficulties seen in autism spectrum condition stem from an inherent departure from typical cognition. The DEP proposes that autistic individuals experience these difficulties mainly in situations where they interact with non-autistic people and that is because there is a breakdown of mutual understanding, empathy and reciprocity. It suggests that this breakdown is caused by the autistic experience being different from that of non-autistic individuals, and this breakdown is in fact bidirectional i.e. non-autistic individuals are faced with similar social and communication difficulties when socializing with autistic individuals. In other words, it suggests that these problems with communication will surface any time there is an incongruity in neurotype, i.e. during autistic – non-autistic interactions, and would be much less pronounced in neurotype-congruent interactions i.e., autistic – autistic or non-autistic – non-autistic interactions. It further advocates that the social communication problems that autistic individuals face are therefore mainly because of having to navigate a largely non-autistic society and not just because of inherent difficulties stemming from autistic cognition. Hence, if the world were autistic, autistic individuals would not have any social and communication difficulties.

If the DEP holds true, the implications go beyond just adjusting our understanding of autistic behaviour and cognition; it has serious consequences for how autistic people are generally treated. For example, it has been suggested that since medical professionals must gauge the emotional and mental state their patients are in, if there is a breakdown of understanding and reciprocity, this can hinder medical professionals’ understanding of their patients’ needs

(Edey et al., 2016). A possible link between autistic individual being misperceived and the increased risk of mental health issues and suicide in the autistic population has also been put forward (Mitchell et al., 2021). Therefore, it is essential that the DEP be further explored and empirically tested.

5.1.2. Empirical support for the DEP

The DEP has garnered empirical support over the last few years. Autistic participants have been seen to have unique communication styles that are dissimilar to those of non-autistic individuals (Heasman & Gillespie, 2019), which are usually not observed since studies predominantly focus on autistic-to-non-autistic mixed interactions.

Some studies have tried to map interactions in neurotype-congruent and -incongruent pairs and groups, and have found that participants have increased preference for those who are neurotype-congruent with them. Chen et al. (2021) observed natural peer interactions and found that autistic students were more likely to interact with autistic peers while non-autistic peers interacted more with other non-autistic peers, and that the purpose and style of interaction differed depending on neurotype-congruency. This preference for neurotype-congruent interaction translates to greater future interest in socialisation (with neurotype-congruent partner) by non-autistic and autistic people alike (Morrison et al., 2020). Both autistic and non-autistic people rate their rapport to be higher in neurotype-congruent interaction versus neurotype-incongruent interactions, which is also corroborated by third-party observers who watched videos of said interactions (Crompton, Sharp, et al., 2020). Autistic participants also reported experiencing an increase in their competence as conversationalists when in dialogue with other autistic strangers, over and beyond even when communicating with people who they were familiar with (G. L. Williams et al., 2021). However, just the presence of a preference, increased rapport, and a self-reported increase in communicative competence do not necessarily equate with better quality and success of communication.

Crompton et al. (2020) empirically tested the effect of neurotype-congruency on communication quality. They used diffusion chains – chains of eight people where the first person heard a story they had to recount to the second person, and the second person had to pass it to the third person, and so on and so forth. They found that stories travelled better in the neurotype-congruent chains: details were retained throughout the autistic-only chains as well as the non-autistic chains and were retained significantly better in neurotype-congruent chains than in the neurotype-incongruent mixed chains, where the rate of decline in detail was much more salient. All participants were aware if the chain they were part of was an autistic chain, a non-autistic chain, or a mixed chain. This underscored that communication was not only preferred in neurotype-congruent groups due to increased rapport and similar communication styles, but also that the quality of communication was better in neurotype-congruent interactions, unlike that in neurotype-incongruent interactions where there was a quantifiable breakdown of communication. This breakdown is even evident in non-verbal communication: motor features that convey affective state of the actor, called 'vitality forms' that are essential to non-verbal communication, were found to be misinterpreted by non-autistic judges when the actors are autistic versus when actors are non-autistic (Casartelli et al., 2020). The question then is if this breakdown is purely at the behavioural level or if neurotype dissimilarity affects higher mental processes as well, in order to investigate what the underlying causes of it are and if it can be mitigated.

For years now, it has been believed that autistic individuals show difficulties in mentalizing – the ability to read mental states of others and infer intention from their mental states (Baron-Cohen et al., 1985; Senju et al., 2009). Therefore, to consider the DEP as a viable explanation for the mentalizing difficulties seen in autism (Rajendran & Mitchell, 2007), it needs to be tested if it also holds true at the cognitive level, in the context of understanding mental states of similar and dissimilar others. Sheppard et al. (2016) found that even when non-autistic participants rated autistic adults in a video as expressive as non-autistic adults when reacting to an event, they were able to correctly interpret the autistic adults' behaviour less often than non-autistic adults' behaviour. However, Sheppard et al. (2016) did not investigate if this was true when autistic adults rated other autistic adults, or when autistic adults rated non-autistic adults. Edey et al. (2016) tested both autistic and non-autistic adults on adapted versions of the Frith-Happé animations (Abell et al., 2000) and found that autistic individuals were better at attributing mental states to animations created by other autistic individuals than those created by other non-autistic individuals. Non-autistic participants were better at attributing the correct mental state to the triangles over all, but in contrast to the autistic group, they showed a greater ability to do so for animations generated by other non-autistic participants versus those generated by other autistic participants.

Therefore, it could be said that not only do autistic people struggle to 'read' the mental states of non-autistic people, but non-autistic people also have difficulties 'reading' autistic people. Since mentalizing or mindreading is thought to be an important component of deception detection (Stewart et al., 2019), does this breakdown in reciprocity translate to deception detection as well? If so, the deception detection ability of autistic receivers for autistic senders should be comparable to that of non-autistic receivers for non-autistic senders. Similarly, non-autistic receivers should struggle to detect lies from autistic senders just as autistic receivers have been found to struggle to detect lies from non-autistic senders, such as those in the Cheating task (see Chapter 4). However, considering that mentalizing, an ability autistic individuals considered to struggle with (Schneider et al., 2013; Senju et al., 2009), is an important cognitive factor effecting deception detection, it is likely that autistic receivers will still struggle to detect lies in comparison to non-autistic receivers, albeit less so when detecting lies from autistic senders versus non-autistic senders.

Therefore, the aim of the study was to test if the DEP holds true in the context of deception detection. I used non-verbal videos of autistic and non-autistic senders lifting metal boxes while they pretended that the boxes were either heavier or lighter than they actually were (pretend lifts) or lifted them as they were without any pretence (real lifts). Participants in this study (receivers) watched these videos and rated all videos on a 5-point deception scale (1 = real; 5 = pretend). Using non-verbal deception videos rather than verbal lies helped mitigate the issue that autistic and non-autistic individuals have different speech patterns (Geelhand et al., 2021; Hubbard et al., 2017; Sharda et al., 2010), and these differences might interact with verbal deception cues. Additionally, past research suggests that people are able to judge whether a person is pretending (i.e., being deceptive) or not when lifting an object (Podda et al., 2017; Schubotz & von Cramon, 2009). On this basis, along with the past literature supporting DEP (Crompton, Ropar, et al., 2020; Edey et al., 2016; Sheppard et al., 2016), I predicted:

H_1: Pretend lifts will be rated higher on the deception scale than real lifts.

H_2: Non-autistic receivers will be better at detecting deception overall, i.e. they will better distinguish between pretend lifts and real lifts than autistic receivers.

H_3: All receivers will be better at detecting deception for neurotype-congruent senders (versus neurotype-incongruent senders). In other words, autistic participants will be better at discriminating between real and pretend lifts for autistic senders (versus non-autistic senders), while non-autistic participants will be better at discriminating between real and pretend lifts for non-autistic senders (versus autistic senders).

5.1.3. The DEP as a type of intergroup bias

The systematic propensity to favour an in-group, i.e. “group of people with shared interests or identity” (‘Oxford English Dictionary’, 2022), versus an out-group, i.e. people who are not part of said in-group, is called an intergroup bias. It means that people have a tendency to evaluate those who are similar to them more positively and those who are dissimilar more negatively or less positively (Hewstone et al., 2002).

Intergroup biases can influence behaviour – actively stereotyping a group, or creating a mental impression of an out-group, which then drives all future interactions with that group. Non-autistic individuals often have a negative first impression of autistic individuals (Alkhaldi et al., 2021; Sasson et al., 2017), who ostensibly come across as appearing and sounding differently to themselves (Faso et al., 2015; Hubbard et al., 2017), and this affects their openness to future interactions with autistic individuals (Morrison et al., 2020; Sasson et al., 2017). Furthermore, these negative first impressions of autistic individuals (on non-autistic individuals) change depending on disclosure of their diagnosis. For example, Sasson and Morrison (2019) found that autistic individuals were rated more positively once non-autistic individuals were made aware of their diagnosis. However, there is also evidence that disclosure of diagnosis can have the opposite effect, i.e. first impression ratings may become even more unfavourable after disclosure, and this often depends on prior attitudes and biases held by the rater instead of the one being rated (Morrison et al., 2019). Intergroup biases affect cognition as well - studies investigating intergroup bias have shown that people are good at understanding mental states of their in-groups, while they struggle to do so for out-groups. Research has shown that this is especially salient in the context of cultural differences Perez-Zapata et al. (R. Adams et al., 2009; Perez-Zapata et al., 2016), but as demonstrated before, this is also present in context of autistic–non-autistic interplay (Edey et al., 2016). This demonstrates that the DEP is likely an extreme case of intergroup bias, where non-autistic individuals hold some preconceived ideas about autistic individuals and this drives their behaviour towards autistic individuals.

5.1.4. Artificially inducing DEP-related breakdown in reciprocity

While cross-cultural designs study intergroup-bias when groups are pre-defined, mentalizing ability has also been studied in context of groups that have been artificially induced, such as when participants have been assigned groups arbitrarily based on trivial criteria (Minimal Group Paradigm; e.g.: Diehl, 1990; Hong & Ratner, 2021). Hackel et al. (2014) categorized participants randomly as either over-estimators or under-estimators following a task where they estimated how many dots flashed on a screen. Participants were then required to assign the presence of a 'mind' (e.g. humans have a mind, robots do not have a mind) on a scale of 1 (definitely has no mind) to 7 (definitely has a mind) to characters on screen that were morphs between humans and computer-generated inanimate faces. Participants showed the need for the characters to be more human-like (than computer-generated) to assign presence of mind when the character was an out-group member versus when the character was an in-group member. Results were also replicated for real-world groups, i.e. labelling the morphs on the basis university groups, rather than over/under-estimator groups (Hackel et al., 2014). McLoughlin and Over (2017) also demonstrated a marked variability in mentalizing based on the characterization of the agent as either in-group or out-group. They showed five and six year old children the Frith-Happé animations (Abell et al., 2000) and told them that some of the triangles were based on other children who were (a) the same or a different gender to them, or (b) from the same or a different geographic location as them. They found that children used more mentalizing language and in greater variety for their in-group in comparison to their out-group.

This shows that intergroup biases do not just manifest when there are ‘real’ groups (e.g. race, religion, gender), but also when people are artificially assigned groups based on arbitrary criteria or are told that stimuli belong to a people belonging to a certain real-world group. This means that intergroup bias can be artificially induced. The question then is if the intergroup biases seen between autistic and non-autistic individuals in the form of DEP can also be artificially created by explicitly labelling individuals as autistic or non-autistic. In

Crompton et al.’s (2020) study participants were aware of which communication chain they were a part of – autistic only, non-autistic only, or mixed chain – and they found that information transfer was more efficient in in-group only chains. In Edey et al.’s (2016) study however, participants were unaware of the neurotype of the animation creator when responding, but mentalizing difficulties for neurotype-incongruent agents were still observed. This poses the question: does this indicate that the intergroup bias stems from inherent differences between autistic and non-autistic behaviours that the participants react to or can it be triggered by artificially creating neurotype-based in-groups and out-groups, indicating that it may instead stem from preconceived notions each group has of the other.

To test the idea of artificially inducing neurotype-based intergroup biases akin to the DEP, Partington et al. (in prep.) used an eye-tracking based false-belief implicit mentalizing paradigm (similar to Southgate et al., 2007), where autistic and non-autistic participants were primed to predict the actions of the agent in the video. Two different agents introduced themselves as either autistic or non-autistic prior to the actual task through audio clips. Although the autistic audio clip was recorded by an actually autistic individual, both agents in the false-belief videos were in reality non-autistic and the agent who was introduced as autistic was counterbalanced across participants. Partington et al.’s (in prep.) preliminary findings show that autistic and non-autistic participants accurately predicted actions for the agent that appeared to be neurotype-congruent, while failing to do so for the agent that was seemingly neurotype-incongruent. Therefore, regardless of the actual stimuli, intergroup bias was observed in mentalizing ability simply by labelling one of the agents as autistic and the other as non-autistic, demonstrating that presence of autistic behaviours may not be required for DEP-related breakdown in reciprocity.

As a second aim of the present study, I wanted to see if the effect of neurotype-congruency expected in deception detection as per the DEP could be synthesised by artificially labelling people as autistic and non-autistic, regardless of their actual neurotype. To investigate this, versions of the Weird Pictures and Two Truths and a Lie deception detection tasks (see chapter 2 for more details) were used – participants watched triplets of videos from multiple senders, and before each trial, they were given information regarding the sender's neurotype. Half of the senders in each of the tasks were randomly labelled as autistic while the other half were labelled as non-autistic. As with previous iterations, participants were asked to guess which of the three statements was a lie for each sender. As mentalizing is thought to be a cognitive factor driving deception detection (Sip et al., 2008; Stewart et al., 2019) and previous studies have shown that people are better at mentalizing about their in-group versus their out-group (Hackel et al., 2014; McLoughlin & Over, 2017) and this is true for fictitious neurotype labelling (Partington et al., in prep.), I expected that:

> **H_4:** Participants will be better at detecting lies from the senders with neurotype-congruent labels, i.e. in-group, in comparison to senders with neurotype-incongruent labels, i.e. out-group. In other words, autistic participants will be more successful in detecting lies for senders who are artificially labelled as autistic (versus those who are labelled as non-autistic) while non-autistic participants will be more successful in detecting lies for senders who are labelled as non-autistic (versus those who are labelled as autistic).

5.2. Method

5.2.1. Participant

Data was collected from adult autistic and non-autistic participants (18 – 60 years) in two waves that used two different recruitment strategies.

In the first wave of recruitment, participants were invited to participate in the Autism@ICN coordinated testing initiative at the UCL Institute of Cognitive of Neuroscience through advertisements on social media (Twitter, Facebook, Instagram, Reddit). People already on the Autism@ICN database were invited to participate via direct emails. Participants were paid the departmental rate of £7.50 / hour (in Amazon UK or Amazon US vouchers) for the background battery they completed as part of the coordinated testing as well as the task battery they undertook specifically for this study. Similar to all prior studies, only fluent English speakers were recruited for the study.

One hundred and three participants were recruited in this wave. However, after employing the same data quality checks as in Chapter 4 (see Chapter 4 for details), only 63 participants were included in the final sample (M_{age} = 30.37 years, SD_{age} = 8.81 years). Eighteen participants were excluded for meeting one or more of the exclusion criteria based on the MaRs-IB (Chierchia et al., 2019), but no participants were excluded for meeting the exclusion criteria based on STW (Baddeley et al., 1993; adapted for online testing as seen in Livingston et al., 2021). Three participants were excluded because they were self-diagnosed as autistic, two were excluded because they were over the age of 60, and 17 were excluded due to incomplete data.

An additional 143 participants were recruited in a second wave of data collection through Prolific (www.prolfic.co). As with the development of deception detection study (Chapter 4), the study was set up to reject participants who did not score at least 60% on the STW. Only one participant who completed the study met the data quality-based exclusion criteria. Data from the remaining 142 participants (M_{age} = 36.45 years, SD_{age} = 9.84 years) were included in the study. All participants were paid the departmental rate of £7.50 per hour through Prolific.

The final sample after all exclusions was made up of 205 participants (103 autistic). All autistic participants confirmed that they had been diagnosed by a qualified clinician; however, due to the online nature of the study it was not possible to confirm the diagnoses of the autistic participants. That said, the autistic participants reported significantly higher autistic traits as measured by the Autism Spectrum Quotient (AQ, Baron-Cohen, Wheelwright, Skinner, et al., 2001). The two groups were comparable on verbal and non-verbal ability, as well as age. Unfortunately, the two groups did not have the same gender breakdown – there were more autistic males than females, while the opposite was true for the non-autistic group. However, since gender was not observed to have an effect on deception detection in Chapter 4 and this gender split is common when recruiting autistic samples due to autistic males being diagnosed at a much higher rate, no further steps were taken to match the two groups on gender. More details about the sample can be found in Table 5.1.

Participants in this study did not take part in the development of deception detection study and vice versa. To increase motivation (as explained in Chapter 2 and 3), participants also had the opportunity to win a bonus in addition to the compensation for their time and effort at the departmental rates – the participant with the best performance in the deception detection tasks in each group was awarded an extra of £20.

Table 5.1.
Sample demographics and group matching

	ASC (*n* = 103)		NT (*n* = 102)		*t*-test
	Mean (*SD*)	Range	Mean (*SD*)	Range	*p*-value
Age (years)	35.27 (*9.93*)	18-57	33.88 (*9.91*)	18-60	.317
Verbal ability[1]	81.23 (*12.95*)	51.67-98.33	83.40 (*9.33*)	60-100	.170
Non-verbal ability[2]	58.21 (*18.95*)	26.58-100	59.61 (*16.25*)	30.65-89.74	.570
Autistic traits[3]	34.11 (*8.94*)	12-50	19.01 (*8.22*)	5-41	<.001
Gender	43 F; 54 M; 6 NB		53 F, 49 M		.026
Handedness	11 L; 84 R; 8 AMB		10 L; 90 R; 2 AMB		.146
L2 English learners	3		15		.003

Note. NB = Non-binary; AMB = Ambidextrous; L2 English learners = Participants who are second language speakers of English.

[1]Verbal ability as measured by accuracy percentage on Spot The Word (STW, Baddeley et al., 1993; adapted for online testing as seen in Livingston et al., 2021). Note that three autistic participants from Wave 1 had an accuracy below 60% but were still included in the sample. This was done because the Wave 1 exclusion criteria was based on above chance performance (50%) as a data quality (attention) check while Wave 2 was set to auto-exclude participants below 60% as the STW in Wave 2 also acted like a pre-screen for verbal ability. As the groups are matched on verbal ability, this was not considered an issue.

[2] Non-verbal ability as measured by accuracy percentage on the Matrix Reasoning Item Bank (MaRs-IB, Chierchia et al., 2019)

[3]Autistic traits as measured by the Autism Spectrum Quotient (AQ, Baron-Cohen, Wheelwright, Skinner, et al., 2001)

5.2.2. Measures

5.2.2.1. Deception Tasks

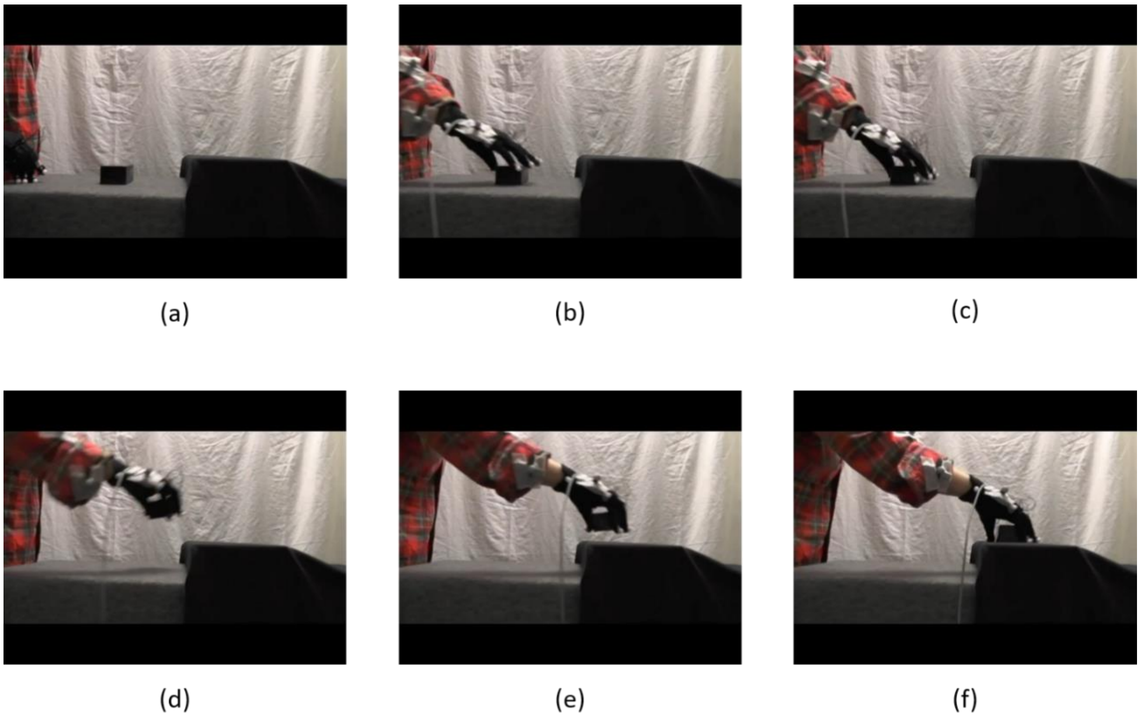
The study used three deception tasks – versions of the Weird Picture, Two Truths and a Lie, and a non-verbal deception task called ‘box lifts’.

5.2.2.1.1. Box lifts – Non-verbal deception task

The Box Lifts task was created using videos recorded by the Social Neuroscience Lab at UCL’s Institute of Cognitive Neuroscience (Hamilton & Kumano, personal communication). During the production stage, 25 autistic and 25 non-autistic age and IQ-matched participants (senders) were invited into the lab to take part in a study investigating movement and kinematic profiles of autistic and non-autistic individuals. Senders were required to wear motion capture gear and were video recorded. As part of the experimental protocol, participants lifted multiple different metal boxes from a table and placed each one on an elevated platform approximately 30cm in front of the table (see Figure 5.1). Although all the boxes appeared identical, they differed in weight – four boxes with Box A being the heaviest and Box D being the lightest. A subset of the trials required senders to lift the box while pretending that the weight of the box was different from reality: they were asked to lift and place a lighter box as if it were heavy and vice versa.

Figure 5.1

Screenshots of a Box Lifts video



Note. Senders in each Box Lifts video started in position (a), went through the motions of lifting a metal box (b-e), and placed the box on the shelf (f).

Videos collected were pre-processed in Matlab and were edited in length based on kinematic cues and eight senders were shortlisted based on the smoothness of their movement data as part of a previous study (Hamilton & Kumano, personal communication). For the purpose of reducing the administration time in this study, six each of autistic and non-autistic senders were randomly selected, and six videos, selected on the basis of the weight of the boxes, were included for each sender:

- 2 x pretend videos: lifting Box A (lightest) as if it was Box D (heaviest), and lifting Box D (heaviest) as if it was Box A (lightest)
- 4 x real videos: 2 Box A real lifts and 2 Box D 'real' lifts, i.e. lifts where the senders were not pretending that the weight of the box was any different than it actually was.

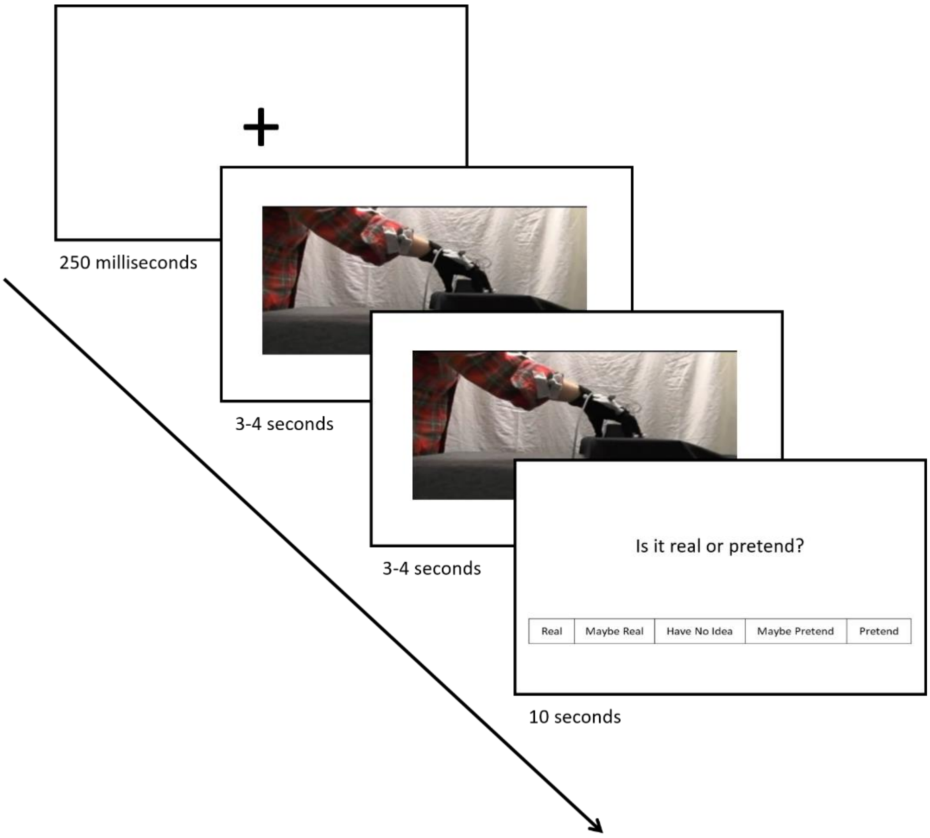
Videos shortlisted for this study were clipped to be three seconds long with an edited one-second freeze added at the start of each video to avoid the possibility that a participant may miss a portion of the video at the start of the trial.

Each trial (Figure 5.2) consisted of a fixation cross for 250 ms followed by one video, which was played twice as the stimuli was so short and the movement was easy to miss. The trial ended with a response screen with a Likert-type rating scale where a response of ‘Real’ was scored 1, ‘Maybe Real’ was scored as 2, ‘Have no Idea’ was scored as 3, ‘Maybe Pretend’ was scored as 4, and ‘Pretend’ was scored as 5. There was a total of 72 experimental trials (6 videos x 12 senders) in the actual task. Trials were arranged in blocks such that all six videos from one sender were blocked together; however within each sender block videos played in a random order. Further, the 12 sender blocks were randomized as well. After every three sender blocks, there was a 10-second break and a screen with an encouraging message, alerting participants to their progress in the task, was displayed: “you are almost halfway through” (after third sender block), “you are halfway through” (after sixth sender block), and “almost at the end now” (after ninth sender block). The whole task took at most 10 minutes to complete.

Participants were told that they would watch several videos, each of which featured either an autistic or a non-autistic individual lifting a small box. They were also made aware that in some of the videos, the actors were instructed to pretend that the box they were lifting was either heavier or lighter than its true weight. Their goal as participants was to judge if each lift was real or pretend. Participants were instructed that, after each video, they would be asked, “Is it (the video) real or pretend?” to which they would have 3 seconds to answer by clicking on the buttons: ‘Real’, ‘Maybe Real’, ‘Have No Idea’, ‘Maybe Pretend’, ‘Pretend’. Participants went through a practice trial consisting of a video from a non-autistic sender (who was not included in the experimental trials) before the actual task.

Figure 5.2

Schematic representation of a trial in the Box Lifts task



Note. Each trial had a fixation cross, followed by one video played twice. At the end of the two playbacks, a response screen with a 5-point Likert-like response buttons was presented.

5.2.2.1.2. Weird Pictures – Double Empathy Problem study (WP-DEP)

The Weird Pictures task (WP-DEP) for this study was created by adapting the WP-3D, used in Chapter 4. The WP-3D task was modified to include some additional elements to help investigate if deception detection was affected by intergroup bias.

Fact Files. Since, the videos in the Weird Pictures were initially shot only with non-autistic senders (see Chapter 2 for details), the WP-DEP only featured non-autistic senders, unlike the Box Lifts. However, in order to investigate if an intergroup bias affects deception detection, half the senders were labelled as autistic. This technique of labelling target stimuli as in-group or out-group is borrowed from minimal group paradigms that are used to study intergroup effects (e.g.: Diehl, 1990; Hong & Ratner, 2021), except that in place of using arbitrary distinctions to create groups, autism diagnosis was used instead. To this end, "*Fact Files"* were created to label senders as either autistic or non-autistic, each consisting of four (fictitious) pieces of information about the sender: Name, age, diagnosis, and a personal fact (such as a hobby or hometown). For example, one of the autistic Fact Files read, "Joshua Coetzer, 19 years old, autistic, enjoys the beach"; and one of the non-autistic Fact Files was "Gerard Jackson, 22 years old, not autistic, from Oxfordshire". The fact files were randomly allocated to the senders for each participant, thus ensuring that no particular sender was always labelled autistic / non-autistic.

Most of the protocol remained the same as WP-3D (Chapter 4). There were four trials and a practice trial. Instead of the videos playing straight away, each trial started with a Fact-File that was displayed for 10 seconds (participants could progress earlier by clicking next). This was then followed by a fixation cross and then a triplet of videos (one lie, one double bluff, and one truth) in a random order. Participants then had five seconds to log their response on the deception-detection response screen by selecting the number corresponding to the description they thought was the lie. This was then followed by a manipulation check.

Manipulation Check. After the deception-detection response screen for every triplet, participants were asked three questions that they answered by sliding the point on a slider:

- How young did the person in the video appear?

  Slider left label: "Young"; Slider right label: "Old"

- How autistic did the person in the video appear?

  Slider left label: "Non-autistic"; Slider right label: "Autistic"

- How happy did the person in the video appear?

  Slider left label: “Unhappy”; Slider right label: “Happy”

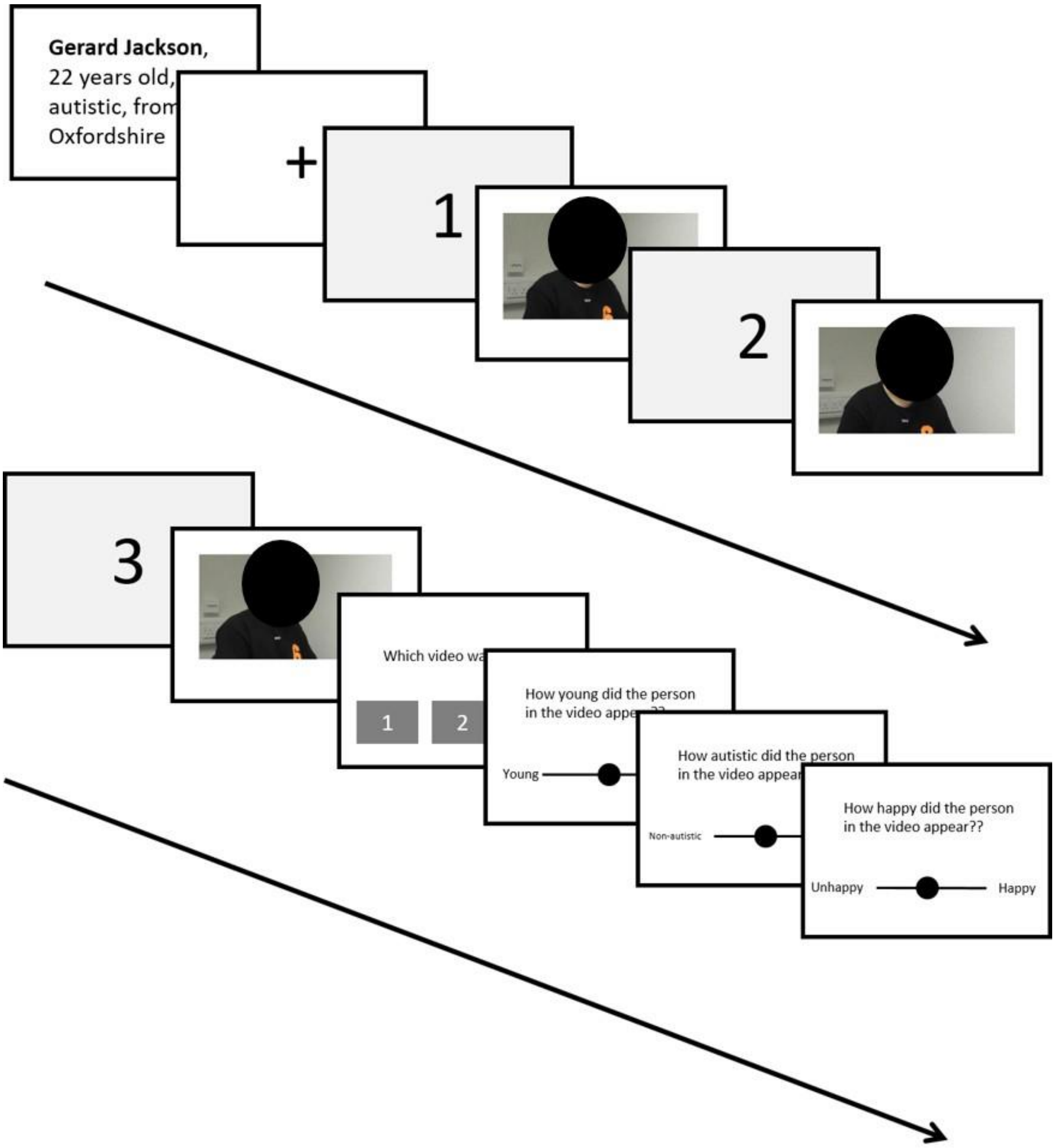
Each question appeared on a separate screen and participants had 10 seconds to respond; a timer appeared in the last three seconds.

If at any point the participants did not respond within the time limit, the task would auto-advance to the next screen / trial. The WP-DEP took about 12 minutes to complete. See Figure 5.3 for a schematic representation of one trial of the WP-DEP.

Instructions. Participants were informed that they would be reading short fact-files for two autistic and two non-autistic people, and that after each fact-file they would watch that person describe some ‘weird’ pictures, which had some unusual things happening in them. The participant was shown an example picture and then was informed that, of the three pictures every person described, for one picture they would describe it truthfully, for one they would double bluff, and for one they would lie. Their (the participant’s) job was to figure out which one was the lie. They were informed that they would have 5 seconds after watching each person’s three descriptions to guess which of three descriptions the lie was, following which they will be asked some general questions about their impressions of the person in the video.

Figure 5.3

Schematic representation of a trial on the WP-DEP



Note. Each trial had a triplet of videos presented in a random order. Before each video a screen with the numbers '1', '2', or '3' was displayed on the screen. At the end of the three videos, a response screen with three response buttons was presented. Faces of senders redacted for the purpose of this report.

5.2.2.1.3. Two Truth and a Lie – Double Empathy Problem study (TTL-DEP)

The Two Truths and a Lie task (TTL-DEP) for this study was created by adapting the TTL-3D, used in Chapter 4. The TTL-3D task was modified to include some additional elements to help investigate if deception detection was affected by intergroup bias.

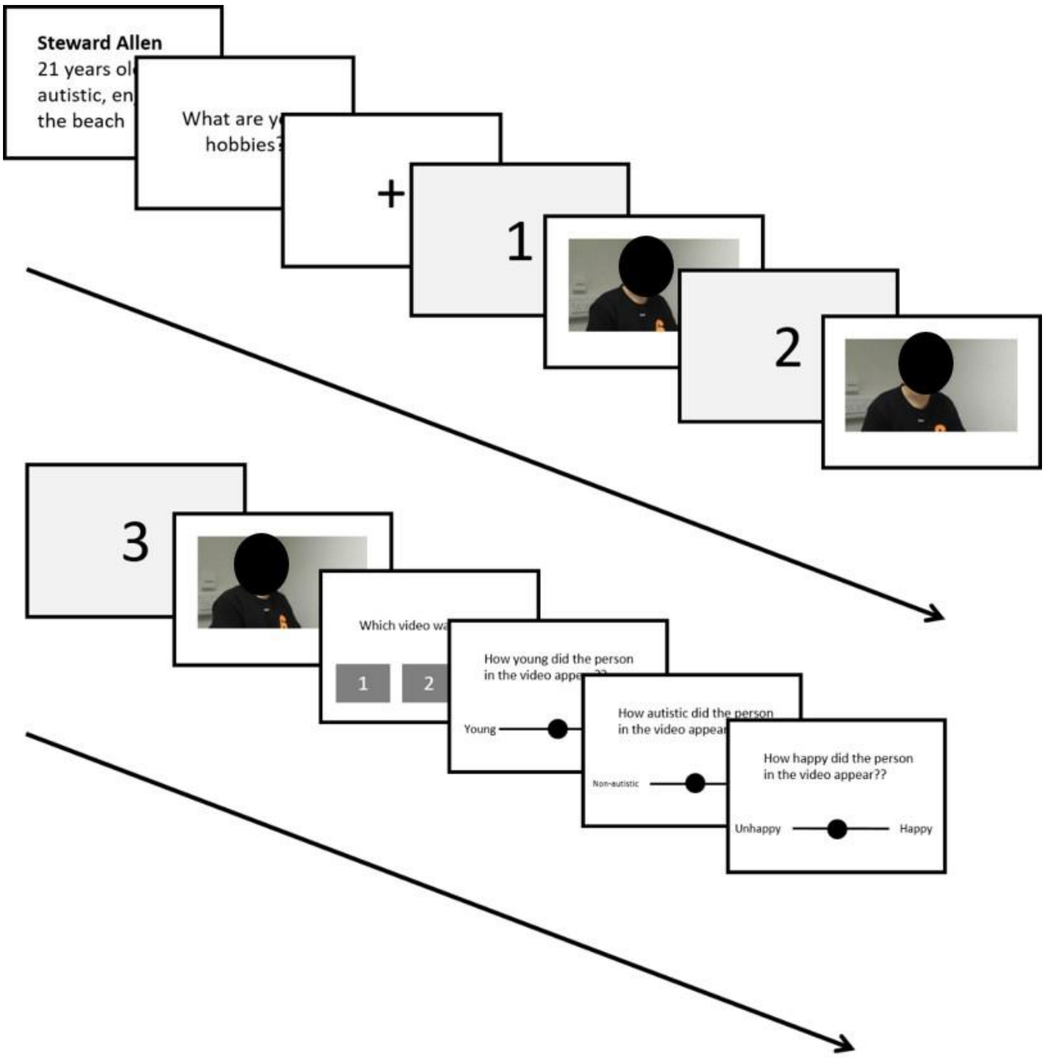
Most of the protocol remained the same as TTL-3D (Chapter 4), except that participants read a Fact-file at the start of every trial, which was displayed for 10 seconds. The Fact-files presented in TTL-DEP were had different names, ages, and personal facts from those used in WP-DEP, but were otherwise presented in the same way and had the same function. This was followed by a screen that displayed the TTL question (i.e. about hobby, summer plans, future goals, or weekend activities) that that specific sender was answering. A fixation cross (250 ms) was displayed before the triplet of videos were shown in random order. After all three videos played, participants had 5 seconds to select which answer they thought was the lie. As with the WP-DEP, the deception-detection response screen was followed by the manipulation check. There were four trials and an additional practice trial.

Participants would automatically be taken to the next screen / trial if they did not respond within the time limit. The TTL-DEP took approximately 12 minutes to complete. A schematic, outlining a trial, is presented in Figure 5.4.

Instructions. As with the WP-DEP, participants in the TTL-DEP were informed that they would be reading short fact-files for two autistic and two non-autistic people and then watch videos of people playing a game called "Two Truths and a Lie" where they answer questions about themselves. For every question, they would watch these people give three answers: two were true and one of them was a lie. Their goal was to guess which one of the three answers was the lie. As with the WP-DEP, participants were instructed on how to respond (selecting one of the buttons corresponding to the video number) and how much time they had after every triplet (5 seconds). Lastly, they were told about additional questions they would be asked about their impression of the people in the video. The TTL-DEP, too, carried a similar disclaimer to the TTL-3D about the videos being recorded pre-pandemic and that the answers the people in the video supplied might not be as relevant in the current climate.

Figure 5.4

Schematic representation of a trial on the TTL-DEP



Note. Each trial had a triplet of videos presented in a random order. The first screen displayed the trial-specific question. Before each video a screen with the numbers ‘1’, ‘2’, or ‘3’ was displayed on the screen. At the end of the three videos, a response screen with three response buttons was presented. Faces of senders redacted for the purpose of this report.

5.2.2.2. Questionnaire measures

5.2.2.2.1. Autistic traits: Autism Spectrum Quotient Questionnaire (AQ)

The Autism Spectrum Quotient (AQ, Baron-Cohen et al., 2001) was used to measure autistic traits. The AQ is a 50-item self-report questionnaire that records participants' responses on a 4-point scale ('slightly agree', 'definitely agree', 'slightly disagree', and 'definitely disagree'). An online version of the questionnaire, available in Gorilla's resource database (www.gorilla.sc), was used for this study.

5.2.2.2.2. Group Identification Questionnaire (Group-IDQ)

As this study investigated intergroup differences, it was essential to make sure that the two groups associated theselves with being either autistic or non-autistic. The Group Identification Questionnaire (Group-IDQ, Doosje et al., 1995) was employed to measure how much the participants identified with the two diagnostic groups. The original questionnaire by Doosje et al. (1995) consists of four statements that are rated on a 7-point Likert scale (1 = 'not at all', 7 = 'very true') to quantify the cognitive, evaluative and affective aspects of the participants' identification to a particular group. It was adapted in this study so that it could record participants' identification with their in-group as well as out-group. Therefore, the Group-IDQ in this study had eight questions – four recording participants' identification with being autistic and four recording their identification with being non-autistic. The eight items in the Group-IDQ were:

- I feel strong ties with **autistic** people.
- I see myself as **autistic**.
- I identify with other **autistic** people.

- I am glad to be **autistic**.
- I see myself as **non-autistic**.
- I identify with other **non-autistic** people.
- I am glad to be **non-autistic**.
- I feel strong ties with **non-autistic** people.

5.2.2.3. Background measures and technical checks

5.2.2.3.1. Verbal ability estimate – Spot the Word (STW)

The Spot the Word task (STW; Baddeley et al., 1993; adapted for online testing by Livingston et al., 2021) used in Chapter 4 was also used in this study to measure participants' verbal ability. It is a word categorisation task where participants are required to speedily and accurately select the actual word in 60 pairs each consisting of one word (e.g. oasis) and one non-word (e.g. broxic). The higher the accuracy percentage, the better the verbal ability.

5.2.2.3.2. Non-verbal reasoning ability estimate – Matrix Reasoning Item Bank (MaRs-IB)

The Matrix Reasoning Item Bank (MaRs-IB; Chierchia et al., 2019) was used to estimate participants' non-verbal ability. As with the development of the deception detection study in Chapter 4, in this study too participants attempted the task for eight minutes, completing as many matrices as they could within that time by choosing the appropriate option by evaluating the relationship between the other shapes in the matrix. Participants' accuracy percentage was calculated out of only the number of items they attempted rather than the maximum number of items in the task (i.e. 80). A high accuracy percentage on the MaRs-IB demonstrated high non-verbal ability, which could in turn be extrapolated as high performance ability.

5.2.2.3.3. Technical checks – bot and sound checks

The sound-checks (James, 2019) employed in Chapter 4 were used at the start of this study too. The sound check doubled as an effective bot-eliminating device as well since bots are unable to fill in free text boxes correctly. Participants had to listen to and type accurately (in a free text box) two words that were played ("octopus" and "chocolate"). Any participants who failed the sound-check three consecutive times were automatically eliminated.

5.2.2.3.4. Surrounding check

Since this was an online study, participants were required to complete a checklist of requirements regarding the environment they were attempting the study in. This was to ensure optimal conditions of participation in order to: a) maintain data quality, and b) avoid any confounds introduced due to the testing environment. All participants were required to confirm that they had fast internet connections that would allow for the videos in the tasks to stream immediately and that they were sitting comfortably and in a quiet place. Participants were advised to wear earphones / headphones and put their mobile phones on silent to avoid distraction.

5.2.3. Procedure

Participants were recruited in two waves – Wave 1 and Wave 2 and the overall procedure and order in which tasks were presented differed between the two waves even though the battery of tasks was the same for both. For both waves, the whole study was created and hosted on Gorilla Experiment Builder (www.gorilla.sc).

Similar to the participants recruited in Wave 1 of the development of deception detection study, participants in Wave 1 of this study were recruited as part of a wider coordinated testing project. A testing session consisted of multiple studies broken down into a number of streams (arranged in parallel) and stages (arranged one after the other) that lasted a maximum of three hours if completed continuously. Participants were, however, not encouraged to do the whole testing session in one go and were encouraged to take breaks between stages.

The background tasks common to all studies in the coordinated testing project, such as those measuring verbal and non-verbal ability and autistic traits, were completed in the very first stage. Depending on the demographics of the incoming participants such as age and the recruitment criteria of any preceding studies, this study was either the second or the third stage of the coordinated testing session.

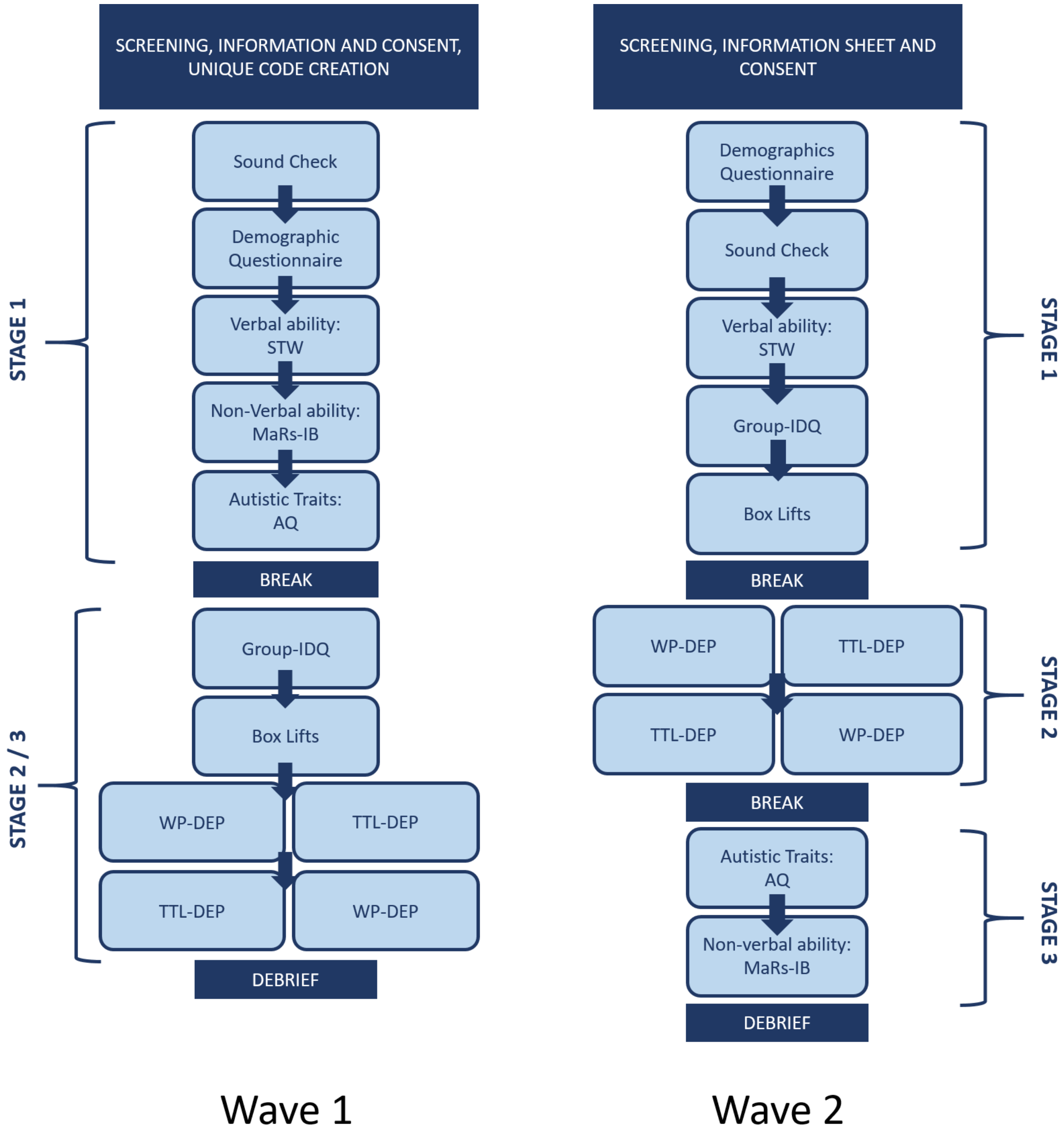
The coordinated session was accessed by clicking on a direct link to Gorilla Experiment Builder (www.gorilla.sc) where they were screened for eligibility, read the information sheet, and gave consent before finally creating a unique participant code and then moving onto Stage 1. Stage 1 (background measures) took approximately 45 minutes, and the tasks exclusive to this study took approximately 40 minutes, making the total time spent on completing all tasks associated with this study less than 90 minutes.

Participants in Wave 2 were recruited specifically for this study via Prolific (www.prolific.co) and were required to complete the battery of tasks in three stages. The first two stages took just under 30 minutes each, and the final stage was shorter – approximately 15 minutes. Participants completed one stage and then waited to be invited into the next stage. Stage 1 started with participants reading the information sheet and filling in the consent form after an eligibility-screening questionnaire.

The chronology of the tasks / questionnaires in both Wave 1 and Wave 2 are presented in Figure 5.5.

Figure 5.5

Schematic representation of the experimental procedure in Wave 1 and Wave 2 recruitment



Note. The study was divided into multiple stages for both Wave 1 and Wave 2. Participants were suggested to take a break between stages in order to reduce fatigue. The order of WP-DEP and TTL-DEP was counterbalanced between participants. Please note that Stage 1 in Wave 1 some participants also completed an additional task after the AQ (viz. a mentalizing measure) that is not part of this study and therefore has not been included in the schematic.

5.2.4. Ethics

All procedures were approved by UCL Research Ethics Committee (Project ID Number: 14807/002). There were no identifiable risks to participants due to this study. All participants read detailed information sheets detailing all risks and benefits of the study.

One concern regarding this study, more specifically regarding WP-DEP and TTL-DEP, was that non-autistic senders' videos recorded during the Lie production stage of this project (see Chapter 2 for details) were labelled as autistic. However, no single individual was singled out to be characterised as autistic and all participants were fully debriefed at the end of this study that none of the senders in those specific videos was actually autistic. Therefore, this adapted use of videos was not considered to be in breach of any ethical code of conduct and indeed this was reviewed and approved by the UCL Research Ethics Board.

The second similar concern was that the decision to label non-autistic individuals as autistic instead of involving actually autistic individuals in the stimuli creation process could be seen as exclusionary. The original research proposal did plan to include autistic individuals and outlined a study to collect original deception stimuli from autistic individuals as that would make the tasks truly equipped to test the Double Empathy Problem. However, videos from autistic individuals could not be collected within the timeframe of this project due to Covid-related restrictions. In order to make sure that the autistic community was not left out of this important process, a (paid) focus group was run with five autistic individuals, recruited via word of mouth from the research groups' social circles but who were unaware of the aims of this research. The focus groups discussed the idea of using the fact-files to mitigate the issue of unavailability of autistic stimuli was discussed. The idea was unanimously approved by the focus group.

Participants who completed the study but who were excluded due to not meeting data quality standards were paid – this was especially relevant for Wave 1 participants. In Wave 2, participants who were auto-rejected due to underperforming on the STW were not paid; however participants were explicitly forewarned about this in the information sheet and assent forms and did not complete the rest of the study.

5.3. Results

5.3.1. Internal consistency of questionnaire measures

The internal consistency reliability of the two questionnaire measures used in this study were examined, determined by the Cronbach's alpha – for the whole sample and also for the groups separately. Cronbach's alpha ranges from 0 to 1, where a higher value indicates higher internal consistency, with 0.70 being deemed as an acceptable level for questionnaires being used in social science research (Kline, 2015).

5.3.1.1. Autism Spectrum Quotient (AQ)

In this study, the AQ was demonstrated to have very good reliability, with a Cronbach's alpha of .93 for the whole sample, .90 for the autistic group, and a more modest but still good .86 for the non-autistic group.

The Cronbach's alpha values for the five subscales are usually found to be lower, sometimes below acceptable levels. The alpha values for the subscales in this study are provided in Table 5.2, along with those found by Hurst et al. (2007) for comparison.

Table 5.2

Cronbach's alpha values for all AQ subscales

Subscale	Across groups	ASC	NT	Hurst et al. (2007)
n	205	103	102	1005
Social	.84	.80	.77	.66
Attention Switching	.80	.73	.69	.41
Attention to detail	.62	.62	.54	.60
Communication	.85	.77	.70	.47
Imagination	.74	.66	.64	.40

Note. Values for Hurst et al. (2007) are reported for comparison

5.3.1.2. Group Identification Questionnaire (Group-IDQ)

The Group-IDQ was made up of four questions each for recording participants' identification with being autistic (GIDQ-ASC) and with being non-autistic (GIDQ-NT). The internal consistency of each of these subscales was calculated separately for each of the two groups. For GIDQ-ASC, the scores showed good internal consistency for the ASC group ($\alpha = .83$), and for NT group ($\alpha = .79$). Similarly, GIDQ-NT too showed good internal consistency for the ASC group ($\alpha = .73$), and for the NT group ($\alpha = .79$).

5.3.2. Group Identification

In order to check that participants (receivers) identified with their own diagnostic group, i.e. autistic or non-autistic, the scores from the Group Identification Questionnaire (Group-IDQ) were analysed. Scores, ranging from 1 to 7 on a Likert scale, across all questions for each of in-group and out-group were averaged for each participant. A mixed 2 (receiver group: ASC / NT) x 2 (group membership: in-group / out-group) Analysis of Variance (ANOVA) demonstrated that receivers identified significantly more with their in-group compared to their out-group; F (1,203) = 375.280, $p < 0.001$, $\eta_p^2 = .649$. However, receiver group also significantly interacted with group membership (F [1, 203] = 16.792, $p < .001$, $\eta_p^2 = .076$): Autistic receivers identified with their in-group significantly less than non-autistic receivers (t [203] = -2.549, $p = .012$, $d = -.356$), while also identifying with their out-group much more than non-autistic receivers did (t [203] = 4.542, $p < .001$, $d = -.635$). Means and standard deviations of how much receivers identified with their in- and out-groups are in table 5.3.

Table 5.3.
Means and standard deviations of Group-IDQ

	All receivers		Autistic receivers		Non-autistic receivers	
	M	*SD*	*M*	*SD*	*M*	*SD*
In-group identification	5.19	1.35	4.96	1.38	5.43	1.28
Out-group identification	2.50	1.11	2.83	1.06	2.16	1.05

Note. M = Mean, *SD* = Standard deviation. A mean score of 1 indicates low identification and a mean score of 7 indicates high identification. A mean score of 3.5 would be a non-biased response.

5.3.3. Investigating the double empathy problem: Box lifts

Participants rated each video on how fake they thought the lift was on a deception scale of 1 (Real) to 5 (Pretend). The ratings across all real and pretend videos and neurotype-congruent and neurotype-incongruent videos were averaged for each participant (Table 5.4). A 2 x 2 x 2 mixed ANOVA was conducted on the deception rating with receiver **group** (ASC / NT) as the between-subject variable, and **authenticity** of box lift (Real / Pretend) and **congruency** between receiver and sender neurotype (Congruent / Incongruent) as the within-subject variables.

5.3.3.1. Main effects

There was no significant effect of group (F [1, 203] = 0.05, p = .823, η_p^2 = .000) or congruency (F [1, 203] = 1.225, p = .270, η_p^2 = .006). Authenticity, however, did show a significant main effect (F [1, 203] = 485.295, p < .001, η_p^2 = .705): participants rated Pretend videos as being more fake than Real videos.

5.3.3.2. Interactions

There was no interaction between Congruency and Authenticity; F (1, 203) = 2.320, p = .129, η_p^2 = .011. Receivers discriminated between real and pretend lifts similarly for in-group and out-group senders.

There was a significant interaction between Congruence and Group; F (1, 203) = 92.736, p < .001, η_p^2 = .314 (Figure 5.6). Post hoc paired t-tests showed that autistic participants rated videos as being more fake in the congruent condition (autistic senders) than in the incongruent condition (non-autistic senders), t (102) = 5.91, p < .001, d = .583. Conversely, non-autistic participants rated videos as being significantly more fake in the incongruent condition (autistic senders) than in the congruent condition (non-autistic senders), t (101) = -7.74, p < .001, d = -.764.

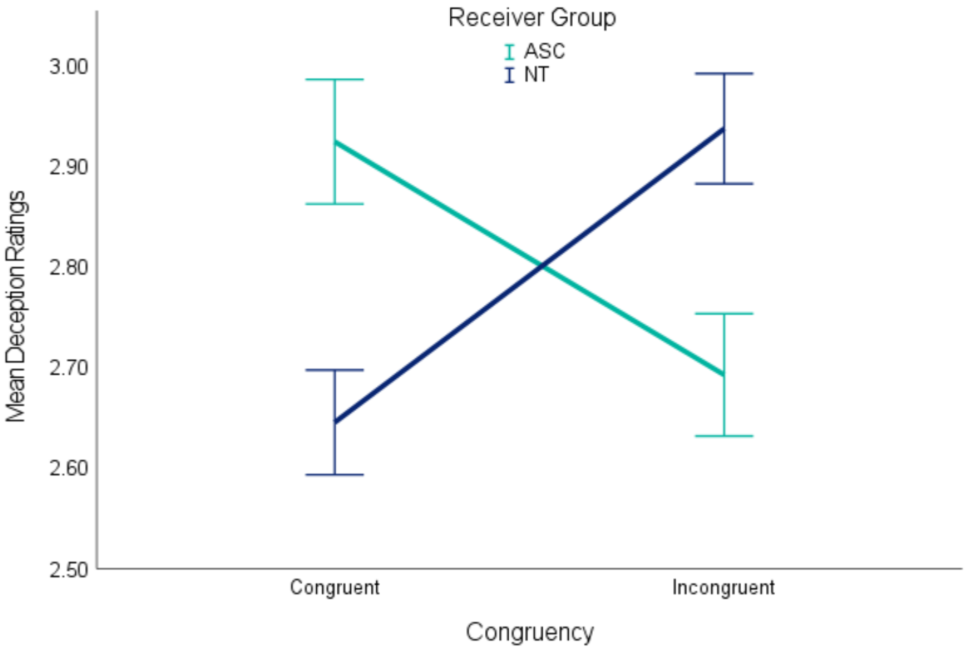
Table 5.4.
Means and standard deviations of video ratings in all conditions and groups

Authenticity	Congruency	Group		
		ASC	NT	Across groups
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
Across Authenticity	Across Congruency	2.69 (0.60)	2.65 (0.51)	2.67 (0.56)
	Congruent	2.93 (0.63)	2.65 (0.53)	2.79 (0.60)
	Incongruent	2.69 (0.62)	2.94 (0.55)	2.81 (0.60)
Real	Across Congruency	2.46 (0.65)	2.37 (0.58)	2.42 (0.62)
	Congruent	2.49 (0.68)	2.28 (0.61)	2.39 (0.65)
	Incongruent	2.43 (0.69)	2.47 (0.64)	2.45 (0.66)
Pretend	Across Congruency	3.16 (0.64)	3.21 (0.54)	3.19 (0.59)
	Congruent	3.36 (0.71)	3.01 (0.59)	3.19 (0.68)
	Incongruent	2.96 (0.69)	3.41 (0.61)	3.18 (0.69)

Note. ASC = autistic receivers, NT = non-autistic receivers, *M* = mean, *SD* = standard deviation

Figure 5.6

Interaction between Congruency and Receiver Group



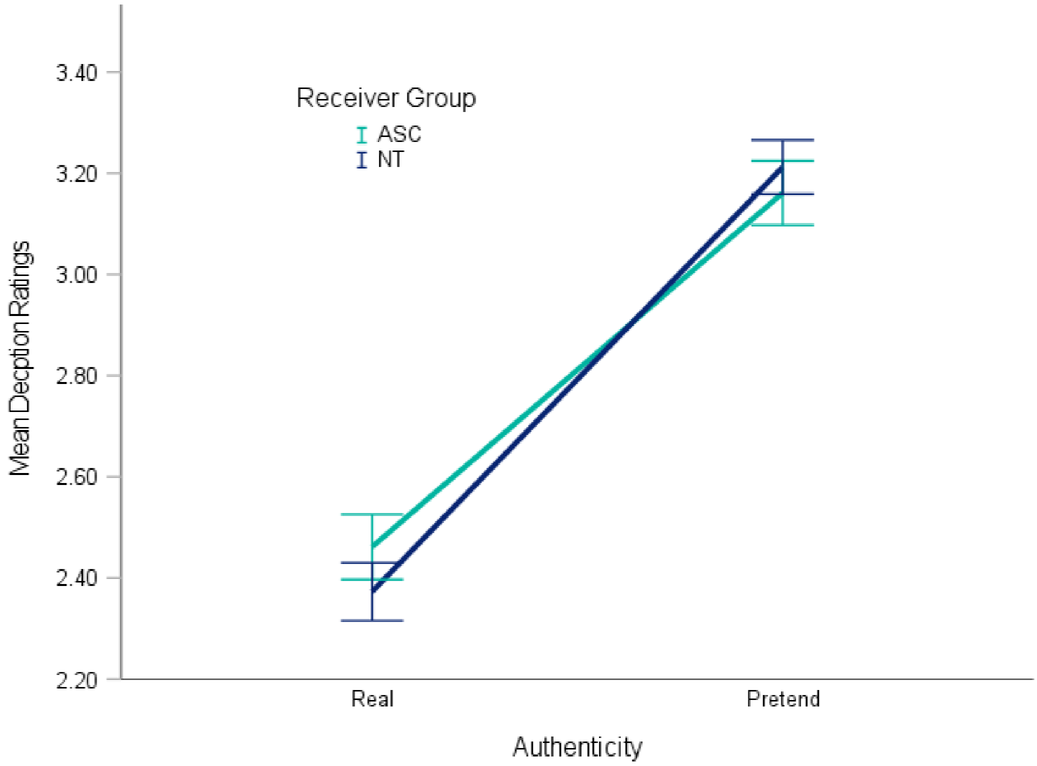
Note. The scale of the graph has been adjusted to illustrate the differences between categories better; the Y-axis does not start at zero. Error bars represent one standard error.

There was a significant interaction between Authenticity and Group; F (1, 203) = 4.249, p = .041, η_p^2 = .021 (Figure 5.7). Post hoc paired *t*-tests showed that both autistic and non-autistic participants rated pretend videos as being significantly more fake than real videos (ASC: t [102] = -13.69, p < .001, d = 1.349; NT: t [101] = -18.00 p < .001, d = 1.782). However, autistic receivers rated real videos as being slightly more fake and pretend videos slightly less fake versus non-autistic receivers, differentiating less between the two types. Indeed, on further investigation it was seen that the difference between the autistic participants' ratings for real and pretend videos (M = 0.70, SD = 0.05) was slightly smaller than that of the non-autistic participants (M = 0.84, SD = 0.47), t (203) = -2.02, p = .044, d = -.283.

There was a significant three-way interaction between Group, Authenticity, and Congruency, F (1, 203) = 40.816, p < .001, η_p^2 = .167.

Figure 5.7

Interaction between Authenticity and Receiver Group



Note. The scale of the graph has been adjusted to illustrate the differences between categories better; the Y-axis does not start at zero. Error bars represent one standard error.

5.3.3.3. Post-hoc ANOVAs

To explore this 3 way interaction further, two separate 2 x 2 repeated measures ANOVAs were conducted, one each for autistic receivers (ASC) and non-autistic receivers (NT), with Congruency (Congruent / Incongruent) and Authenticity (Real / Pretend) as the within subjects factor. The analysis was split by group in order to able to test if deception detection performance (difference between ratings in real and pretend lifts i.e. authenticity) was different depending on congruency between the neurotype of senders and receivers. Furthermore, the significant interaction between congruency and group indicates that congruency affects the two groups differently – thus the interactions between congruency and authenticity may vary by group as well.

5.3.3.3.1. Congruency and deception detection in autistic receivers

For the autistic receivers, there was significant main effect of Authenticity; F (1, 102) = 180.807, p < .001, η_p^2 = .639. Autistic receivers rated pretend lifts as being significantly more fake than real lifts, indicating intact ability to detect deception. There was also significant main effect of congruency, F (1, 102) = 35.001, p < .001, η_p^2 = .255. Surprisingly, autistic receivers rated their in-group senders (i.e. autistic senders) much higher on deception, i.e. rated them as being more fake across trials, than their out-group senders (i.e. non-autistic senders), regardless of their actual veracity.

There was a significant interaction between Authenticity and Congruency for autistic receivers; F (1, 102) = 35.801, p < .001, η_p^2 = .260 (Figure 5.8a). Paired t-tests showed that autistic receivers rated senders in both congruent and incongruent conditions comparably in the real lifts (t [102] = 1.501 p = .136, d = -.148), but they rated in-group senders more fake than out-group senders for the pretend lifts (t [102] = 7.298, p < .001, d = 716). This showed that autistic receivers were better at detecting deception in the in-group than the out-group but were more likely to identify other autistic individuals as being deceptive.

5.3.3.3.2. Congruency and deception detection in non-autistic receivers

The ANOVA for non-autistic receivers presented a significant main effect of Authenticity; F (1, 101) = 324.136, p < .001, η_p^2 = .762. Participants rated pretend videos significant higher on deception than real videos. There was also a main effect of congruency, F (1, 101) = 59.953, p < .001, η_p^2 = .372: the non-autistic group rated out-group (i.e. autistic senders) videos much higher on deception in comparison to the in-group (i.e. non-autistic senders).

There was a significant interaction between Authenticity and Congruency for the non-autistic receivers, F (1, 101) = 10.493, p = .002, η_p^2 = .094 (Figure 5.8b). Post-hoc paired t-tests revealed that the non-autistic group rated out-group senders as being more fake than in-group senders for both real lifts (t [101] = -4.092 $p < .001$, d = -.405) and pretend lifts (t [101] = -7.428, $p < .001$, d = -.735). However the difference between incongruent and congruent was significantly more pronounced for the pretend lifts (M = -.40, SD = 0.54) than real lifts (M = -.19, SD = 0.46); t (102) = 3.239, p = .002, d = .321. This indicates that, contrary to what was predicted, non-autistic receivers were better at deception detection for the out-group than they were for the in-group.

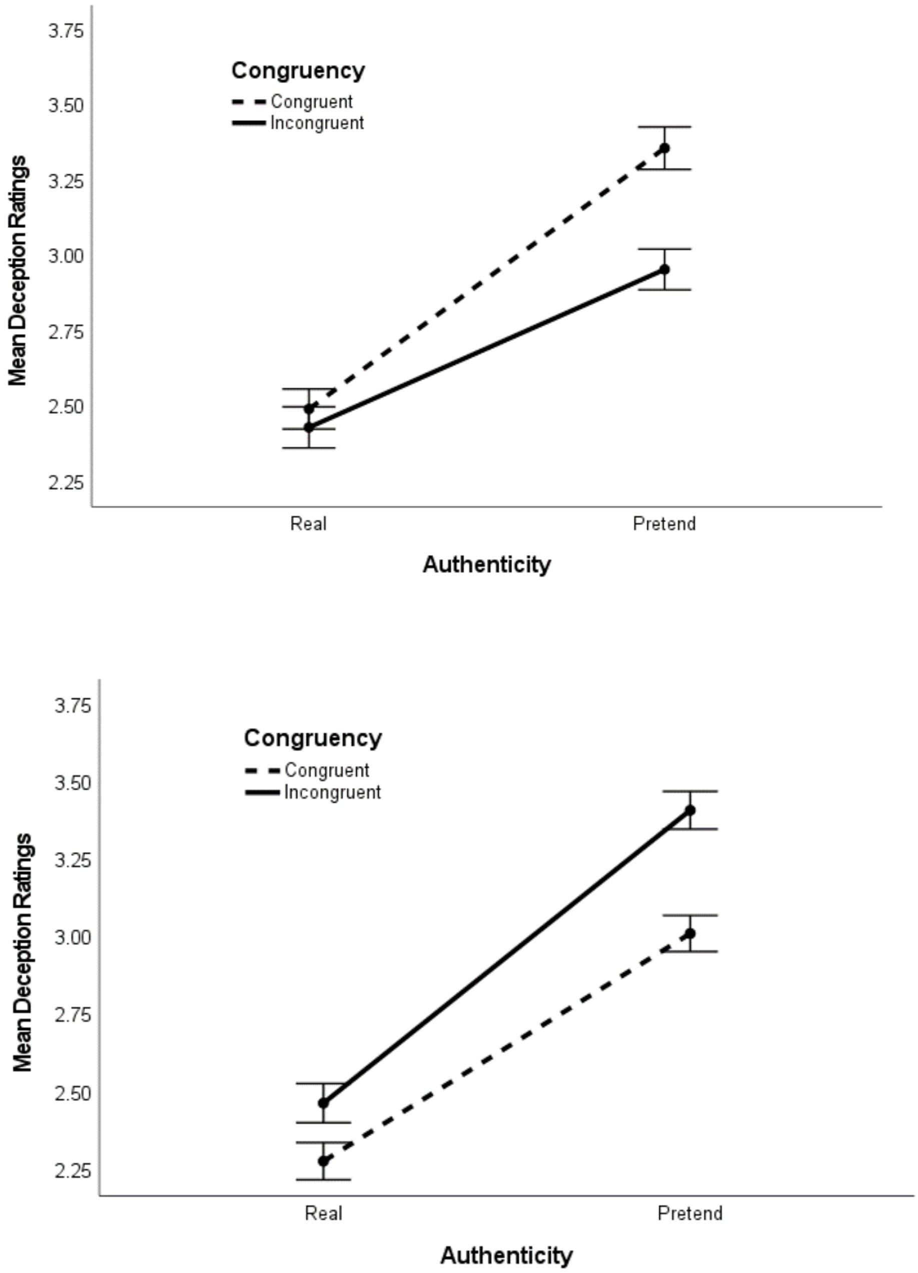
(Rest of page intentionally left blank)

Figure 5.8

Interactions between Authenticity and Congruency for (a) ASC and (b) NT receivers

(a)

(b)



Note. The scale of the graphs has been adjusted to illustrate the differences between categories better; the Y-axis does not start at zero. Both (a) and (b) are on the same scale. Error bars represent one standard error.

5.3.4. Investigating induced intergroup bias: WP-DEP and TTL-DEP

To investigate if the groups were better at detecting lies from senders they believed to be in-group members and were worse at detecting lies from senders they thought to be out-group members, ideally a 2 x 2 mixed ANOVA, with congruency as the within subjects factor and receiver diagnosis as the between subjects factor, would have been run. However, considering that that data from WP-DEP and TTL-DEP was ordinal (possible scores of 0, 1, and 2 for each sender label), an ANOVA was unsuitable. Unfortunately, a direct non-parametric equivalent to the mixed ANOVA has not been developed, as per my knowledge. To circumvent this issue, non-parametric equivalents of t-tests, akin to post-hoc tests of a mixed-ANOVA, were run individually. While unconventional, this method of using multiple non-parametric comparisons has been used before, where the data violated the assumptions of parametric tests (e.g.: Nelson et al., 2017).

For what would have been the two main effects, a between subjects Mann-Whitney U-test was employed to test the main effect of receiver group, and a within subjects Wilcoxon signed-rank test was used to test the main effect of congruency. To test the interaction between the two factors, a between subjects Mann-Whitney U between ASC and NT receivers on the difference between their congruent and incongruent score. To avoid type 1 errors due to the increased number of comparisons, a more stringent alpha value of .01 was used to ascertain significance.

5.3.4.1. Manipulation checks

As neither the WP-DEP nor the TTL-DEP had actual autistic senders and used ‘fact-files’ to influence participants’ perception of the senders’ diagnosis, it was imperative to check if participants believed the manipulation. Therefore, participants were asked to rate the senders on how autistic they thought the sender was after each trial. Although the slider that the participants used to respond to the manipulation check question did not have numbers on the ends (instead had ‘non-autistic’ and ‘autistic’), their responses were recorded on a scale of 0 (non-autistic) to 100 (autistic).

The manipulation was successful in both tasks. In the WP-DEP, participants rated Autistic-labelled (ASC-L) senders (M = 53.27, SD = 18.81) significantly higher on autistic appearance versus Non-Autistic labelled (NT-L) senders (M = 37.74, SD = 18.23); $t(204) = 8.06$ $p < .001$, $d = .563$. Similarly, in TTL-DEP too ASC-L senders (M = 55.30, SD = 20.10) were rated higher on autistic appearance than NT-L senders (M = 40.87, SD = 18.60) as well; $t(204) = 8.12$ $p < .001$, $d = .567$.

5.3.4.2. Performance against chance

As this was the first time the novel tasks in their current form – with sender neurotype labels - were being used as part of an experimental paradigm, it was important to check that the task was working as expected, to be able to draw conclusive inferences. To do so, firstly, like the Chapter 2 and Chapter 4, it was assumed that for the task to be deemed to be working, non-autistic (NT) participants would need to score above chance. By design, NT receivers expected to perform worse in the incongruent condition i.e. for ASC-L senders. Therefore, only the NT receivers’ performance on the NT-L condition were tested against chance to gauge if the tests were working.

One-sample Wilcoxon Signed Rank tests were performed against a test value of 0.67 (chance: 2 trials x 1/3 probability of getting the correct answer by chance). NT receivers scored significantly above chance for the TTL-DEP, but their performance was not significantly above chance for WP-DEP (Table 5.5).

Table 5.5

Non-autistic receivers' performance against chance on congruent condition on novel tasks

	n	*M* (*SD*)	*Mdn*	*T*	*z*	*p*	*r*
WP-3D	102	0.86 (0.65)	1	3136.00	1.75	.081	0.17
TTL-3D	102	1.04 (0.76)	1	3687.00	3.60	<.001	0.36

Note. WP-DEP = Weird Pictures (Double Empathy Problem), TTL-DEP = Two Truths and a Lie (Double Empathy Problem).

Considering that non-autistic receivers were unable to perform above chance on the WP-DEP, the findings from WP-DEP were considered to be uninterpretable as the task was deemed to be not working. Therefore, the results from the WP-DEP are not included below and can instead be found in Appendix B. Only the analysis on the data collected on the TTL-DEP is reported and interpreted here.

5.3.4.3. Main effect of receiver diagnosis

A Mann-Whitney U-test was run to check if the receiver groups differed on how well they judged lies in the TTL-DEP, regardless of the sender labels. Participants could obtain a score between 0 to 4. Autistic receivers (M = 1.76, SD = 1.00, Mdn = 2) were slightly worse at detecting deception than non-autistic receivers (M = 2.09, SD = 1.14, Mdn = 2) were. Although this difference was significant at .05 level, it was not so at the chosen alpha level of .01; U = 4912.00, z = -0.841, p = .401, r = -.05

5.3.4.4. Main effect of congruency

A Wilcoxon signed-rank test was run to see if lie detection differed depending on congruency between sender label and receiver diagnosis. As there were only 2 NT-L trials and 2 ASC-L trials, participants could achieve a score between 0-2 for each of the congruency conditions.

There was no significant effect of congruency in the TTL-DEP, i.e. no difference in lie detection between the congruent condition (M = 0.91, SD = 0.76, Mdn = 1) and the incongruent condition (M = 1.01, SD = 0.71, Mdn = 1), T = 4488.50, z = 1.448, p = .147, r = .101.

5.3.4.5. Interaction between receiver diagnosis and congruency

To examine if congruency between sender labels and receiver diagnosis affected performance differently for autistic and non-autistic receivers, an 'interaction score' was calculated by subtracting the receivers' scores in incongruent conditions from their scores in congruent condition. This score could range from -2 to +2 and a higher score denoted better lie detection for congruent senders. A significant difference between groups in their interaction scores would indicate that congruency affected lie detection performance differently in each group.

A Mann-Whitney U-test demonstrated that autistic receivers (M = -0.20, SD = 1.06, Mdn = 0) and non-autistic receivers (M = 0.01, SD = 0.93, Mdn = 0) did not significantly differ in how congruency affected their performance in the TTL-DEP, U = 4717.00, z = -1.324, p = .186, r = -.09.

These results must be interpreted in the context of the main effect of congruency. As there was no effect of congruency – i.e. across groups, receivers did not show any difference in their deception detection abilities in either congruent or congruent conditions – the interaction demonstrates that neither group performed better in any condition. In other words, the results from the TTL-DEP show that both groups performed comparably in all conditions.

5.4. Discussion

In this chapter, I aimed to examine if the double empathy problem (DEP) plays a role in how autistic and non-autistic individuals detect deception. Autistic and non-autistic individuals watched videos of autistic and non-autistic senders lifting metal boxes, in some of which the senders faked how heavy the box actually was. It was found that, as expected, receivers were able to discriminate between pretend and real box lifts (H_1), and while both autistic and non-autistic receivers rated pretend lifts as significantly more fake than real lifts, non-autistic receivers were slightly better at discriminating between real and pretend lifts (H_2). This shows that autistic individuals, while sensitive to deception to a degree, are not as good at detecting deception as their non-autistic counterparts. As predicted, autistic receivers detected deception by in-group senders better than they did for out-group senders. However, contrary to predictions, non-autistic receivers detected deception with significantly more effectiveness for out-group than in-group senders (H_3).

Another aim of this chapter was to investigate if the proposed effects of the Double Empathy Problem on deception detection could be manufactured artificially by inducing intergroup bias in autistic and non-autistic receivers by merely labelling half the non-autistic senders as autistic instead of using actually autistic senders. Receivers were asked to watch videos of senders (labelled as either autistic or non-autistic) attempting to deceive about a picture or about themselves and had to pick out which of their statements were lies. Contrary to what was expected, in spite of believing the senders' autistic / non-autistic labels, receivers' deception detection was comparable for both senders who had neurotype-congruent labels and those who had neurotype-incongruent labels (H_4). In fact, there was no difference in deception detection ability between the two groups as measured by the novel tasks, unlike the box lifts.

5.4.1. Double empathy and deception detection – Box lifts

To study if the double empathy problem (DEP) – the idea that the social communicative difficulties that autistic people face is not just due to ‘autistic cognition’ but because of a mutual breakdown of understanding and reciprocity (D. E. M. Milton, 2012) – a non-verbal deception paradigm was used. Both autistic and non-autistic senders were video recorded while they lifted boxes of various weights, and in some lifts, pretended that the box weighed more or less than it actually did. Autistic and non-autistic receivers in this study watched and rated each of these videos on how real or pretend they thought the box lifts were. As per the DEP, it was expected that receivers would be better at detecting deception from the neurotype-congruent senders versus neurotype-incongruent senders. In other words, autistic receivers would detect deception better for autistic senders and non-autistic receivers would be better at detecting deception from non-autistic senders.

This prediction was partially supported: autistic receivers were better at discriminating between real and pretend lifts for their in-group, i.e. autistic senders. This aligns with the limited literature that empirically tests the DEP. Similar to Crompton et al. (2020), who found that autistic-autistic communication was much more successful than autistic–non-autistic communication, the results here can be interpreted as autistic receivers being more adept at picking up deception cues from other autistic people, although unlike Crompton et al.’s (2020) study, receivers in this study were unaware of the diagnosis of the senders. Considering that mentalizing is thought to be an essential factor affecting deception (Stewart et al., 2019), it is unsurprising that the current study particularly supports the findings of Edey et al (2016), who studied the possible effects of DEP on mentalizing. Edey et al (2016) found autistic people were better at attributing the mental states of animations created by other autistic individuals, similar to how autistic receivers in the current study detected deception better for autistic senders. They also found that non-autistic participants were better at mentalizing overall, akin to how the non-autistic receivers in this study showed superior deception detection abilities in general.

However, the findings for the non-autistic receivers in this study deviate from Edey et al.'s (2016): unlike in the past study where non-autistic participants were more adept at mentalizing for other non-autistic people, the non-autistic receivers here do not show greater skill in detecting deception for non-autistic senders. This also goes against Sheppard et al. (2016), who found that non-autistic participants were better at interpreting the behaviour of other non-autistic adults versus autistic adults. In fact, non-autistic receivers in the current study discriminated between real and pretend lifts much better for their out-groups, i.e. autistic senders, rather than their in-group, i.e. non-autistic senders. In other words, they showed better deception detection for the autistic senders, their out-group

Although this contradicts the limited past work on neurotype based intergroup bias in mentalizing, the non-autistic receivers' propensity to detect lies by out-group more efficiently than those by the in-group is in agreement with sparse literature on intergroup deception where groups were divided on criteria other than neurotype. Fan et al. (2022) found that people were more sensitive towards lies told by those who had an opposing attitude towards mask mandates. They attribute this increased ability to detect lies for the out-group to having an increased truth-bias towards their in-group. This has been substantiated by Clementson (2018) as well, who found that people exhibited a bigger truth bias for political candidates aligned with their own party. Dunbar (2017) suggests that the increased ability to discern deception in the outgroups stems from the out-group members deviation from what is considered as normal communication patterns, which leads to the receivers viewing the out-group sender as less trustworthy. This includes both verbal cues that may be considered atypical, such as accents, or non-verbal cues, such as different gestures and expressions. People also detect lies better and judge people more harshly when senders are from different cultures (Fan et al., 2022; Rozmann & Nahari, 2021) and races (Lloyd et al., 2017). Therefore, the non-autistic receivers' performance can be explained by our previous understanding of intergroup deception.

It has been proposed that intergroup bias stems from learnt social norms and how biases towards in-groups and out-groups may differ depending on the social situation (Saarinen et al., 2021). From a mentalizing point of view – it has been suggested that it is beneficial to mentalize more for an out-group if they want to harm you, in order to be able to predict their behaviour (Molenberghs & Louis, 2018). It is plausible, then, that an increased sensitivity to the out-group occurs when there is an antisocial element present – such as deception or another form of insincerity. For example, when asked to discriminate fake smiles from real smiles, participants showed a heightened sensitivity to the fake smiles from the out-group than from their in-group, possibly because participants are wary of the out-group and their motives (S. G. Young, 2017). This differential intergroup bias towards in-groups and outgroups depending on different social situations – specifically, a possibly heightened bias towards outgroup in antisocial situations could also explain why the results for the non-autistic receivers here differs from that from Edey et al. (2016) or Crompton et al. (2020). The tasks in these past studies did not have an antisocial element and therefore mentalizing or communication was not more increased for out-groups (versus in-group), unlike this current study. Given that deception has an inherently negative connotation socially, it is therefore understandable that in this study, where participants were aware that they would be lied to, and in past intergroup deception studies, participants had a stronger bias towards detecting deception in the out-group.

But then again, if intergroup deception proposes better detection for out-group members, this calls into question why autistic receivers in this study did not show an elevated deception detection ability for their out-group, i.e. the non-autistic senders. This could have a few potential reasons. First, how much an individual identifies with their in-group has been found to be a major predictor of the extent of intergroup bias at play (Aberson & Howanski, 2002; Ahmed, 2007). In the present study, even though both autistic and non-autistic receivers in this study identified with their in-group more than their out-group, autistic receivers showed less adherence to group identity and, importantly, reported significantly less identification with the in-group than non-autistic receivers reported. This, therefore, could be a possible reason as to why they do not show the otherwise observed trends of intergroup deception. Second, autistic individuals are much more likely to be exposed to other non-autistic individuals than non-autistic individuals are to autistic people. As contact with out-group has been found to significantly reduce intergroup bias (Lemmer & Wagner, 2015; Pettigrew & Tropp, 2006), this could be another reason why autistic individuals do not show better deception detection capabilities for out-group unlike non-autistic individuals. That said, the study did not gather data on past experiences with the out-group, therefore this explanation needs more direct testing.

Lower adherence to in-group identity and more exposure to out-group may justify why autistic receivers did not show heightened deception detection for non-autistic senders, but in the light of past literature on intergroup deception, does not explain why autistic receivers instead showed significantly better detection for autistic senders. Support from previous DEP literature on mentalizing notwithstanding, apart from going against the findings of previous studies investigating intergroup deception, the results from autistic receivers additionally contradicts the inclinations shown by non-autistic receivers in this study. The two neurotypes therefore seem to demonstrate distinctly different trends when it comes to intergroup deception. Possibly then, the best way to explain the conflicting results for the two neurotypes in this study lies in the fact that both autistic and non-autistic receivers showed better deception detection for autistic senders. In other words, regardless of their own neurotype, receivers detected deception from autistic senders much more efficiently than from non-autistic senders. This could be because autistic senders have been found to be poor at deception (Baron-Cohen, 1992; A. S. Li et al., 2011; J. Russell et al., 1991) and the autistic senders in this paradigm were less able to pretend about the weights and leaked more cues, thus making them easier to judge. Alternatively, it could be because of the perceived credibility of the sender: deception detection is better for senders who are perceived to be less credible (George et al., 2014) and autistic people have been found to be perceived as lacking credibility (Lim et al., 2022).

It is interesting to note, however, that receivers of both neurotype not only detected deceit by autistic senders with more efficiency, but they also rated autistic senders as more deceptive regardless of the senders' actual veracity or the receivers' own neurotype. This imitates what Lim et al. (2022) found: that autistic people are misjudged as being deceptive even when they are telling the truth; and further builds on their work by introducing a condition where autistic individuals were actually deceptive. This general tendency to rate autistic senders as more deceptive is also in line with past accounts of non-autistic participants judging autistic subjects as less favourable (Alkhaldi et al., 2021; Sasson et al., 2017) and generally having a negative implicit attitude towards autistic individuals.

On the other hand, the findings here contradict autistic people's explicit attitudes towards other autistic people, extrapolated from their general preference of autistic individuals for conversation and socialising (Chen et al., 2021; Crompton, Sharp, et al., 2020; Morrison et al., 2020; G. L. Williams et al., 2021). Autistic people's implicit attitudes towards other autistic individuals have not been previously recorded, as far as I am aware. Some past intergroup research on race, religion, and caste (social hierarchy in Hindus) has shown that minorities may show a favouritism towards an outgroup majority (Dasgupta, 2004; Dunham et al., 2014; Newheiser et al., 2014; Newheiser & Olson, 2012) and this could explain autistic receivers' more favourable ratings for the non-autistic senders (compared to autistic in-group). However, these studies suggest that an in-group favours a majority out-group when the out-group is higher on the social ladder and therefore, this reasoning needs to be explored further with some caution, as it is not currently understood if autistic and non-autistic individuals perceive there to be any social status discrepancy between the two groups. Future studies could investigate implicit social status perceptions alongside intergroup bias amongst autistic and non-autistic individuals to explore if this line of investigation has merit.

Bearing in mind that the receivers in this study were merely instructed to rate how real or pretend the box lifts were, without any indication of which specific lifts were performed by autistic senders, this paradigm could be taken as an implicit measure of the autistic receivers' attitudes towards other autistic individuals. This is therefore the first study, to my knowledge, that proposes that there is a possibility that autistic people may perceive other autistic people in a more negative light than previously believed.

5.4.2. Artificially induced neurotype-based intergroup bias in deception

The second aim of this study was to test if the proposed effects of DEP on deception could be synthesized in paradigms where the groups were artificially created, to investigate if the DEP captured a breakdown that was linked to some fundamental differences between autistic and non-autistic behaviour or if it could be explained by group identities and preconceived notions. The second part of this study was designed to probe if the effects of the DEP, i.e. an increased ability to detect deception from a neurotype-congruent sender could be replicated when the sender was merely labelled as an out-group or in-group, rather than actually belonging to that group.

To this end, the novel deception tasks created specifically for this thesis (Chapter 2) were adapted to include labels – fact files – that informed the receivers of the senders' neurotype. However, unlike Chapter 4, the WP-DEP in this did not work as non-autistic participants were found to be performing at chance. This unfortunately made any results derived from this task uninterpretable, and all further inferences drawn are based only on the TTL-DEP.

The senders used in the TTL-DEP here were the same as used in the TTL in previous chapters, and therefore were actually non-autistic. Despite believing the labels, i.e. believing that half of the senders were autistic, receivers of either neurotype did not show any difference in deception detection for either neurotype label. In other words, receivers were equally adept at detecting lies from those who were labelled as the out-group as they were for those who were labelled as the in-group.

Interestingly, the groups did not differ in deception detection generally on the TTL-DEP, regardless of the sender labels. This contradicts the results from the box lifts, where there was a group difference in how well receivers discriminated between pretend and real lifts. The findings do partially replicate the results from the novel task in Chapter 4 – performance on TTL-DEP does not significantly differ between the groups.

While hardly any studies have attempted such artificially induced groupings for neurotypes, the findings here go against the one study that has attempted to investigate mentalizing by artificially labelling non-autistic actors as autistic. Partington et al. (in prep.) found that autistic and non-autistic participants were worse at attributing mental states to the actors who were labelled as their out-group than to actors who were labelled as belonging to their in-group. However, it is important to note that their paradigm used a more elaborate 'labelling' – participants heard an audio clip where the subjects introduce themselves, the autistic version of which was recorded by someone who was actually autistic. Therefore, there is a possibility that the intergroup bias observed by Partington et al (in prep.) was actually triggered by verbal and vocalic related cues as both autistic and non-autistic individuals have been found to be sensitive to prosodical characteristics in autistic speech (Hubbard et al., 2017). The current study used simple 'fact files' that were essentially made of a line of text informing the receiver of the senders' diagnosis along with other brief and irrelevant biographical information such as age. Therefore, the group-inducing methods used in this study were closer to that used by other studies investigating artificially induced intergroup bias in mentalizing (e.g.: Hackel et al., 2014; McLoughlin & Over, 2017) and tested the possible effects of labelling and any preconceived notions / biases without the influence of the differences between autistic and non-autistic behaviour.

The findings from the TTL-DEP therefore do not align with either the heightened deception detection for in-group, predicted from the past intergroup mentalizing literature, or the elevated deception detection for the out-group, suggested by the past intergroup deception literature. Despite receivers believing the 'fact-file' manipulation, they did not show better deception detection ability for the autistic (labelled) senders in the TTL-DEP to match what was observed in the box lifts task. This could suggest that the intergroup biases observed when autistic and non-autistic receivers judge autistic and non-autistic senders cannot be manufactured artificially. It may indicate that inherent differences in behaviours such as speech (Geelhand et al., 2021), movement (J. L. Cook et al., 2013), and eye-movement (Madipakkam et al., 2017) etc. cause a gulf between autistic and non-autistic senders and receivers that cannot be replicated by mere labelling, unlike other minimal-group and real-world group paradigms using studies which are based more on group-identification.

However, caution must be exercised when drawing conclusions from the TTL-DEP. Firstly, the TTL-DEP is a novel task, without much investigation into its validity and reliability. Also, there were only two trials per condition and that limited how sensitive the tasks were to capture the variance in deception detection ability. Secondly, and more crucially, the inferences drawn here are based on a *lack of difference* between groups, which for obvious reason makes any speculations less conclusive. Therefore, while it is possible that DEP-related intergroup bias may only stem from inherent difference in autistic and non-autistic behaviour and cannot be induced by mere labelling, the opposite is also possible if the tasks used here to test the concept were not sensitive enough to pick up the intergroup bias.

5.4.3. Limitations and future directions

This study is not without its limitation. As with the development of deception detection study (Chapter 4), this study also took place online, and as such there was no way to confirm participants' diagnosis beyond the information they provided. The drawbacks of using a self-report measure over other diagnostic tools such as the Autism Diagnostic Observation Schedule (ADOS; Gotham et al., 2007), and using the AQ specifically instead of other self-report questionnaires such as the Social Responsiveness Scale for Adults (SRS-A; Constantino et al., 2003) mentioned in previous chapters are also relevant here.

One of the reasons for choosing the Box Lifts videos for this study was that they were non-verbal and that mitigated the differences between autistic and non-autistic speech (Geelhand et al., 2021; Hubbard et al., 2017; Sharda et al., 2010). However, autistic individuals also have distinctive movement profiles (J. L. Cook et al., 2013) that may have contributed to autistic senders being perceived with more suspicion by non-autistic receivers, as deviation in non-verbal cues can be misinterpreted as deception cues (Dunbar, 2017). That said, bearing in mind that these atypicalities in movement are part of the autistic profile outside the laboratory and play a role in any deception they produce or others detect in real life, it is counterproductive to eliminate their effects in a deception paradigm studying intergroup deception. Future studies may attempt to parse the effect that distinctive autistic movement may have on how deceptive autistic senders are perceived to be by collecting data on senders' motor profile. In a similar vein, future studies could replicate results from this non-verbal deception paradigm by using verbal deception paradigms, while investigating the moderating effects of atypical speech patterns.

Receivers showed better deception detection for autistic senders when senders were actually autistic and were not explicitly labelled as such (Box lifts) but not when senders were labelled as autistic and were actually not (TTL-DEP). One alternative reason why receivers did not show a similar intergroup pattern of deception detection in TTL-DEP as the Box Lifts could be the explicit neurotype labelling. Negative attitudes that non-autistic individuals have towards autistic people may reduce when people are labelled as autistic (Maras et al., 2019; Sasson & Morrison, 2019). Therefore, the very fact that the senders were explicitly labelled – in order to induce intergroup bias – could have in fact resulted in a less negative attitudes towards them, an increase in credibility, and therefore reduced deception detection for that group. However, the labels could have changed the perceptions for the worse as well – as demonstrated by Morrison et al. (2019) and the direction of change may be dependent on prior existing prejudices the receivers have.

Therefore, in the future, prior attitudes towards autistic individuals need to be measured to gauge how they may interact with explicit labelling and intergroup bias. Additionally, paradigms should be created where both autistic and non-autistic receivers judge both autistic and non-autistic senders either when they are explicitly labelled or when they are not labelled, giving us a better control for the effects of labelling for as well to draw a conclusive understanding of whether intergroup biases based on neurotype can be artificially manufactured. It was, in fact, a part of the initial study design to include autistic senders as well, however due to the CoViD-19 pandemic, stimuli from autistic individuals could not be collected in time to include as part of the paradigm.

5.4.4. Supplementary findings based on a sample defined by AQ cut-offs

As with the chapter on development of deception detection (Chapter 4), the analyses for this chapter were rerun in order to increase the confidence in the inferences that can be drawn from the results. The same criteria by Woodbury-Smith et al. (2005) used in Chapter 4 were used here as well: autistic participants who scored below 26 and non-autistic participants who score above 32 were excluded. The results for these additional analyses can be found in Appendix C.

The newer sample, henceforth called the AQ-defined sample, saw considerable reduction in sample size for the autistic group: about 25% of the autistic participants in the original sample were excluded. For the non-autistic group, this was less drastic with a loss of only 5% of the original sample. As with the previous chapter, the AQ-defined autistic group showed an increase in verbal and non-verbal ability over the original sample, and the mean AQ scores for the autistic and non-autistic group increased and decreased respectively, which was expected.

The resultant analyses had outcomes that were, for the most part, unchanged from the original sample's. The seminal finding that autistic receivers were thought to be more deceptive than non-autistic receivers generally and also when they were not actually being deceptive held for the AQ-defined sample as well. The only dissimilarity between the results found originally and those found for this sample was that autistic receivers differentiated between Real and Pretend lifts as well as non-autistic receivers. In other words, the ability of distinguishing between authentic (truth) and fake (deception) behaviour, i.e. to detect deception, was comparable between the groups. While a departure from the findings from the original sample where the autistic group were distinguished between Real and Pretend lifts to a slightly lesser degree than the non-autistic group, this aligns with the finding from the AQ-defined sample in Chapter 4 where the groups performed comparably on the deception detection tasks.

Given that a similar loss of significant group differences was observed in Chapter 4, the possible explanations for it there are also valid here. The potential loss of power is a concern for the analyses here as well since a fourth of the autistic sample was lost during the exclusion process. And since here too the AQ is used to define the groups, the concerns raised in the previous chapter regarding the potential disadvantages of this process also hold true for this sample and these analyses.

However, the alternative idea presented previously that these group differences found in the original sample may have become non-significant in the AQ-defined sample due to a reduction of noise in the sample might also be relevant here. This finds some support as the group identification scores in the AQ-defined sample are examined against those from the original sample. The AQ-defined autistic group identified with their in-group slightly more and with their out-group slightly less than they did in the original sample. This could indicate that by excluding participants from the autistic group based on the AQ, it may have led to excluding participants who were either not genuinely autistic or may not identify with being autistic as strongly, which could suggest that the findings in the AQ-defined sample may be more valid.

Nevertheless, the findings of this study relied heavily on the receivers’ group identification, more so than their actual diagnosis. In other words, in order to test the double empathy problem, which could be regarded as an autism specific intergroup bias, the extent to which the participants identified with the either being autistic or not was more important than whether they were actually autistic or not. Therefore, it can be argued that, apart from drawing conclusion on the potential difference in deception detection ability, the exclusion of participants on the basis of AQ may not have been as constructive as it likely was in the previous chapter, and that the more conducive variable for grouping here would have been participants' group identification. Also, it bears reminding that the strongest findings of the study are more reflective of the senders and not the receivers, which is again something that is unlikely to have been affected by the AQ-defining of the receiver sample.

Therefore, it may be more prudent to proceed with some caution when interpreting the findings from the AQ-defined sample as the usefulness of such a method of synthesizing the groups could be questioned. At the same time the possibility that the groups may not be significantly different in their ability to detect deception should be considered but with the understanding that such results are derived from analyses that are likely less powerful than the original analyses.

5.5. Conclusions

There are multiple takeaways from this study. First, it re-establishes that adult autistic individuals can detect lies successfully, albeit marginally less successfully than non-autistic individuals. Second, we see that the DEP might have some merit – while autistic individuals show signs of favouring their in-group when it comes to deception detection, i.e. they are able to distinguish deception by their in-group better, the same is not true for non-autistic receivers. But this too could be taken as support for the idea that non-autistic individuals struggle to understand autistic behaviour – their suspicion due to different-from-norm movement profiles of autistic senders and, subsequently, a possible 'othering' of autistic individuals could be a reason as to why non-autistic individuals show a heightened ability to detect deception in autistic individuals.

Third, the effect of neurotype-congruency on deception detection could not be synthesized artificially by labelling non-autistic senders as autistic in this study, possibly suggesting that the intergroup biases seen in DEP might stem from inherent differences in autistic and non-autistic behaviour and mere labelling in the absence of these behaviours likely cannot trigger them. However, it is highly probable that the lack of intergroup bias found due to labelling is because the novel tasks were not sensitive enough to pick up these effects.

Finally, an unexpected finding of the study was that autistic individuals, regardless of whether they are deceiving or not, are thought to be more deceptive. Considering that this may be due to fundamentally different characteristics of autistic individuals and not just intergroup dynamics or group identity has far-reaching consequences. This includes how autistic individuals are perceived in their day-to-day lives such as in interpersonal relationships, in situations that affect their wellbeing such being doubted in a medical Situation, or in high stakes situations such as in criminal investigations - when they are witnesses, or even more gravely, when they are suspects.

Chapter 6. General Discussion

In this thesis, I aimed to add to our understanding of the factors that affect deception – how it interacts with mentalizing, how it develops with age, how it is impacted in autism, how it is influenced by group membership. First, in Chapter 2, I elucidated the issues with traditional deception production and detection paradigms, and sought to create novel tasks that attempted to rectify these issues. Alongside collecting stimuli for a better deception detection paradigm, I examined various factors affecting deception production in Chapter 3. Chapter 4, focussed on the central questions of my thesis – how mentalizing and deception develop with age in autistic and non-autistic individuals and what their impact on peer-victimization and mental health is. Finally, in Chapter 5 I explored if autistic and non-autistic adults were better at detecting deception from those who shared their neurotype and if a neurotype intergroup bias based on deception detection could be artificially manufactured in the absence of true neurotype diversity.

In this final chapter, I start by summarising the key findings from each of the three experimental chapters (3, 4, and 5). Subsequently, I draw on the findings from these chapters to address two central themes: deception and autism. First, I summarise what this thesis adds to our understanding of factors that affect deception. Second, I deliberate on what this study tells us about deception detection in autism specifically. Third, I expound on what implications this has for autistic individuals. In the final sections of this chapter, I talk about some of the overarching limitations of my doctoral project and future directions.

6.1. Summary of experimental chapters

In Chapter 3, I examined how cognitive effort affected deception production – specifically how senders chose between two types of deception that are thought to have different associated cognitive loads and how it affected their success on two tasks that presumably required different amounts of cognitive effort. I tested 42 non-autistic 18-24 years old males on the two novel deception production tasks I created in Chapter 2 – the Weird Pictures (WP) and the Two Truths and a Lie (TTL). For the Weird Pictures task, senders could either chose to lie or double bluff to deceive the receiver. Contrary to what was expected, senders did not show a preference for double bluff, which was thought to be the less cognitively effortful option. Secondly, senders were more successful at deceiving on the TTL than on the WP, even though the TTL was thought to be the more cognitively strenuous of the two by the virtue of being more personal, and thus emotionally charged, than describing pictures. The other half of Chapter 3 involved mapping the associations between deception production success and the various factors that are thought to affect it. The results did not support the hypotheses and deception production was not found to relate to either mentalizing, autistic traits, or demeanour.

Chapter 4 focused on the development of mentalizing and deception detection from late childhood, through adolescence, to early adulthood in autistic and non-autistic individuals. Results revealed that mentalizing ability was only marginally different between the groups, with the autistic group being slightly weaker. Furthermore, mentalizing ability developed with age only in the non-autistic group but not in the autistic group. Deception detection also mimicked the trends in mentalizing: autistic individuals were weaker at deception detection in comparison to their non-autistic counter parts, and that deception detection developed with age only in the non-autistic group. While mentalizing ability did not predict any variance in deception detection ability for the autistic group, it was a marginally significant predictor for the non-autistic group. In Chapter 4, I also investigated the impact of poor deception detection abilities and found that deception detection significantly predicted both peer-victimization (although only marginally in the non-autistic group) and mental health, the effect on the latter mediated by the former.

Finally, in Chapter 5, I explored alternative explanations for the relatively weaker deception detection performance of autistic receivers. It was hypothesised that, in line with the double empathy problem, autistic receivers would be better at detecting deception from autistic senders and weaker at detecting deception from non-autistic senders, but also non-autistic senders would be weaker at detecting deception from autistic senders while being better at detecting deception from non-autistic senders. It was found that both autistic and non-autistic receivers were better at detecting deception from autistic senders than non-autistic senders. In fact, autistic senders were rated much higher than non-autistic senders were on deceptiveness regardless of whether they were actually deceiving or not. A similar result was not found when I attempted to replicate these results in tasks where all senders were non-autistic but some were labelled as being autistic, leading to the tentative conclusion that neurotype-based intergroup bias in the context of deception detection cannot be manufactured artificially.

6.2. What does this project tell us about deception?

Over the three experimental chapters, I have tried to profile the abilities and characteristics that drive individual differences in deception ability. It is important to consider that deception is a social activity and as such, detection and production are two sides of the same coin since deception always requires two people – the sender, a person who is deceiving, and the receiver, the one who is being deceived. The two processes are complementary – deception production success means deception detection failure, and vice versa, regardless of whether the two processes happen at the same time or asynchronously, for example, the baseline study in Chapter 2, which was essentially a deception detection task informed the deception production abilities that were analysed as part of Chapter 3. As a result, any inferences that I draw are keeping this complementary relationship in mind.

6.2.1. Relationship between spotting lies (deception) and reading minds (mentalizing)

From the results of Chapter 3 and 4 it is observed that while production does not seem to be related to mentalizing, detection appears to be weakly related to mentalizing. Here, I attempt to extrapolate what the nature of the relationship between mentalizing and deception is and how this relationship may be stronger than is observable from the direct associations tested in the earlier chapter. I do so by drawing inferences from their common relationships with other variables, such as autistic traits.

Mentalizing ability is found to vary with autistic traits such that low autistic traits coincides with higher mentalizing ability; however, this relationship is more prominent for individuals who report low autistic traits than who report high autistic traits (Mintah & Parlow, 2018; Murray et al., 2017; Nijhof et al., 2017). Findings from Chapter 4 demonstrated that deception detection similarly was negatively predicted by autistic traits but only in the non-autistic sample, not in the autistic sample. In other words, for the non-autistic sample only, those who reported higher traits performed worse on the deception detection tasks, but this variability was not observed in the autistic sample who performed similarly within their group regardless of their autistic symptomology. I propose that this may indicate a possibility that deception detection, at minimum, is affected by mentalizing, and at best, is an indicator of mentalizing itself.

Deception production in Chapter 3 does not demonstrate the same relationship with autistic traits in non-autistic adults – deception production ability does not weaken with an increase in reported autistic traits. This could mean that deception *production* potentially does not have a cognitive profile influenced by mentalizing. An alternative explanation can be extrapolated from the fact that individuals with high autistic symptomatology do not necessarily struggle with explicit mentalizing (Bowler, 1992; Ozonoff et al., 1991; Peterson et al., 2007). This could mean that the reason that there was no observable relationship between autistic traits and deception production is because deception *production* emulates a more *explicit* form of mentalizing. However, caution must be exercised when drawing conclusions from the absence of an observable relationship.

While deception production could potentially be emulating explicit mentalizing, parallels can be drawn between deception detection and implicit mentalizing. Autistic individuals show difficulties in implicit mentalizing while non-autistic individuals do not (Schneider et al., 2013) – a pattern observed in deception detection abilities in Chapter 4 as well. Therefore, it could be speculated that while deception *production* requires a more *explicit* mentalizing, deception *detection* requires a more *implicit* form of mentalizing. Logically, this fits what we know of deception theoretically: deceiving someone requires intent by definition (Zuckerman et al., 1981), and that itself leads it to be an explicit process, as it requires conscious effort and thought. The sender needs to make an effort to understand what the mental state of the receiver is – what they know, believe, feel – and then actively instil a false belief in them. On the other hand, deciphering if someone is being dishonest is intuitive – in real life, except for situations where the receiver suspects the sender of some wrongdoing, people do not have prior knowledge that someone may be lying to them. Therefore, it is understandable that it may tap into more implicit, unconscious processes.

It is however important to consider that some of these findings were not present when the samples in Chapter 4 and 5 were refined on the basis of autistic traits. Since it is not certain if this was because of a reduction of power in the supplementary findings or because the effects in the original findings are due to noisy data, further investigation using multiple explicit and implicit mentalizing tasks in samples with definitive autism diagnosis is required to substantiate these claims. The patterns emerging from this project are encouraging, and while not conclusive, indicate that mentalizing and deception are related, adding to a growing body of literature (Sip et al., 2008; Stewart et al., 2019; Talwar & Lee, 2008).

6.2.2. Influence of sender characteristics on deception success

Law et al. (2018) found that it was the characteristics of the sender, and not the attributes of the receiver that influence the outcome of any deception. Sender demeanour, a measure of how believable / credible / honest a sender comes across as, regardless of their veracity, is found to be one of the major factors explaining variance in deception judgements (C. F. Bond & DePaulo, 2008; T. R. Levine, 2010). Results from Chapter 3, however, demonstrated that sender demeanour did not correlate with how successful they were at producing deception. A point to note though is that the novel tasks created in Chapter 2 were designed to attempt to detect individual differences in receivers' deception detection abilities without demeanour confounding them, and as such, demeanour was kept constant across all veracity types. This could explain why the effect of demeanour found in past studies was not observed in this particular study. For that reason, it is not conclusive evidence challenging past research i.e. the absence of a relationship in this study alone cannot be taken as evidence that sender demeanour does not contribute to variance in deception success.

The idea that "deception judgements depend more on the liar than the judge" (C. F. Bond & DePaulo, 2008, p. 489) finds unexpected support in Chapter 5. I set out to investigate the double empathy problem in the context of deception detection, that is, if deception success improved if the sender was neurotype-congruent with the receiver. The results revealed that regardless of receiver neurotype, receivers were better at distinguishing pretend videos from real videos from autistic senders. This was true even when the sample was further refined to increase confidence in the neurotype groupings by ensuring that the groups had autistic traits at a level usually expected of their respective groups. In other words, the senders' neurotype was what drove deception judgement and not the neurotype incongruency in the receivers. One possible reason for this is that autistic individuals have been found to be poor at deception (Baron-Cohen, 1992; A. S. Li et al., 2011; J. Russell et al., 1991) or alternatively because autistic individuals are in general perceived to have less credibility (Lim et al., 2022). However, considering that these explanations too highlight characteristics of the senders, this only provides more support to the suggestion that deception detection, at least in this instance, was driven by sender characteristics.

The fact that receivers showed increased deception detection for one group of senders over another is even more remarkable when viewed in context of the stimuli that was used in Chapter 5. Senders were judged on very short non-verbal clips (3 seconds) that only showed their arm picking up a box. This left very little room for personalisation of the recordings i.e. there was limited scope for any difference between each individual video that could have affected deception judgement, which was not related to the actual sender. Therefore, significantly better deception detection for one group of senders over another from the scant information available makes it all the more probable that deception judgement is influenced by sender characteristics.

6.2.3. The novel tasks

In chapter 2, I created two novel tasks, the Weird Pictures (WP) and the Two Truths and a Lie (TTL), that attempted to address the limitations of past deception studies. The strengths and limitations of the tasks were discussed in detail in Chapter 2. Here I comment on what can be gleaned from their use in the empirical chapters.

6.2.3.1. The effect of content of deception

In Chapter 3, senders were more successful in producing deception in the TTL compared to WP; in other words, the independent receivers found it much easier to detect lies in the WP than in the TTL, thus leading to the conclusion that senders leaked more cues due to the content of the deception in WP than in TTL. Contrarily, in Chapter 4, a much larger sample of receivers than that in Chapter 3 appeared to be better at detecting lies in TTL than WP, judging by how well receivers performed against chance for both tasks. Results from Chapter 4 also replicate trends found in Chapter 2 (Pilot 2) – that can be similarly deduced from receivers' performance against chance – and is more in line with the prediction that the emotional content in the form of autobiographical information will make lies told in the TTL easier to judge. Therefore, it is possible that the small sample size in chapter 3 (every sender was judged by only 17-21 receivers) led to a type 2 error. One significant point that needs to be considered is that the WP and TTL trials shortlisted for Chapter 2 (pilot-2) and subsequently used in Chapter 4 were supposedly the most transparent, or in other words, the easiest to judge of the larger set of stimuli used in Chapter 3. Therefore, it is not known how the videos from non-transparent senders in either task would fare when judged by a bigger sample of receivers. Therefore, the full range of WP and TTL triplets, like in Chapter 3, will benefit from being judged by a bigger sample of receivers before conclusions regarding the detectability of deception based on their content can be drawn.

6.2.3.2. The validity of the tasks

The TTL was the task that correlated with the Cheating Task in Chapter 4 – a task that is well established as a deception detection task in the literature. This is perhaps unsurprising however, considering that senders in both were deceiving about their own self, even though the Cheating task had much higher stakes. The WP on the other hand involved deception about illustrations that the senders saw for the first time. Therefore, it is possible that the deception detection abilities measured by the WP may be largely affected by other factors, such as receivers' imagination, generativity or even the ability to suspend disbelief, considering that the 'weird' pictures consisted of atypical scenarios (such as a boy walking his tortoise on a leash). This could also be a reason why the WP was found to be more sensitive than the TTL in detecting differences between autistic and non-autistic individuals (Chapter 4) as autistic people are known to also have difficulties in these additional cognitive skills (Craig & Baron-Cohen, 1999; Miller & Bugnariu, 2016; Samson & Hegenloh, 2010). More thorough testing, with the full range of WP and TTL videos that allows for the variance in performance to be captured more efficiently, and where receivers' abilities such as imagination and generativity are recorded and controlled for, will contribute to our understanding of how exactly these tasks work and how valid they are as measures of deception detection.

6.2.3.3.Effectiveness of the tasks

One of the strengths of the novel tasks over other traditional deception detection paradigms was that demeanour was kept constant across veracities, in order to measure individual differences in receivers without the confounding effects of sender demeanour. However, whether or not the tasks were successful in this regard needs to be assessed in the future to gauge the tasks' effectiveness in this regard. This can only be tested when the same stimuli are judged by receivers in a condition where demeanour is constant – i.e. truths and lies are from the same sender compared with a condition where demeanour is not constant – i.e. truths and lies are from different senders. If demeanour predicts deception success in one condition and not the other, it will lend support to the idea that keeping demeanour constant during deception judgement gives a more effective measure of receiver characteristics.

6.3. What does this project tell us about deception detection in autistic individuals?

6.3.1. Can autistic individuals detect deception?

Over chapter 4 and 5, deception detection ability in autistic individuals was tested on four different tasks, each with their own strengths and weaknesses. An unfailing finding across all tasks and studies was that autistic individuals are indeed able to detect deception.

In Chapter 4, this was evident in their above chance performance in the transparent condition of the Cheating Task, the Weird Pictures task (WP) and the Two Truth and a Lie task (TTL). In Chapter 5, this could be gleaned from the fact that they were capable of successfully distinguishing between pretend lifts and real lifts.

6.3.2. Are autistic individuals as good deception detectors as non-autistic individuals?

The question then is how autistic individuals' deception detection ability compares with the deception detection ability of non-autistic individuals. Over the four tasks that have compared deception detection by autistic and non-autistic individuals, three have demonstrated group differences – the Cheating task, the Box lifts, and the WP-3D, suggesting that autistic individuals, while able to detect deception above chance in general, are not *as capable* of detecting deceit as non-autistic individuals. On the other hand, outcomes of one of the four tasks – the TTL-3D – did not demonstrate autistic individuals' deception detection capability to be any weaker than non-autistic individuals'. Thus, it is important to speculate as to why the different tasks may have generated contradictory findings before any conclusion can be drawn about autistic people's skill in deception detection relative to non-autistic people's deception detection skill.

One potential reason for this discrepancy between tasks could be sender demeanour. As mentioned earlier, sender demeanour is considered to be one of the major predictors of variance in deception judgements (C. F. Bond & DePaulo, 2008; T. R. Levine, 2010). The findings from both Williams et al. (2018) and the Cheating task in Chapter 4 show that non-autistic participants are better at detecting deception in the transparent condition and perform slightly worse than autistic participants in the non-transparent condition (albeit this difference is non-significant). An interpretation of this could be that non-autistic individuals rely heavily on demeanour to make their deception judgements, while autistic individuals do not. Put simply, in the transparent condition, where the senders' demeanours match their veracity, the non-autistic receivers may use this to their advantage to detect deception, while autistic receivers do not and consequently underperform. It would be sensible to then assume that any other deception detection paradigm where demeanour is not controlled for will give the non-autistic individuals a significant advantage as sender demeanour will help them detect the deception.

Although the Box Lifts task had multiple videos from the same sender, the receiver had to make a truth-lie judgement for each individual video. This means that non-autistic individuals were still potentially able to use demeanour to decide which lifts were fake and which were not, leading them to be much better at detecting deception than autistic receivers in the same task. The WP and the TTL on the other hand, were designed to control for demeanour across senders, thus reducing the influence of demeanour on receivers' deception judgements. This is potentially the reason why autistic participants performed as well as non-autistic participants in TTL-3D: non-autistic participants could no longer rely on demeanour to pick the lie out of the three videos they were shown, making their deception skills comparable to that of autistic participants who relied less on demeanour to detect deception to begin with.

Another factor that could have led to the divergent outcomes of the different deception detection tasks could be related to the truth bias. Truth bias is a receiver's propensity for assuming that the sender is being truthful regardless of whether their actual authenticity. In lab studies, where receivers are aware that some of the stimuli they will experience is going to be deceptive, an ability that would help them be more successful in deception detection would be inhibiting their own truth bias. This is more so in the case of task designs where the response takes the form of deciding if the stimuli is truth / real or a lie / fake, such as in the Cheating Task or the Box Lifts. If not inhibiting one's truth bias, a receiver would be predisposed to categorising more stimuli as truths than lies, including deceptive stimuli. On the other hand, inhibiting truth bias may not be so important for a task that by design forces the receiver to choose the deceptive stimuli from a set of videos. If the only way to respond was to identify the lie amongst a set of videos, such as in the WP or TTL, the truth bias may not play such a big role, as the receiver is not being asked *if* a video is a lie, but *which* video is the lie, making the presence of the lie more definitive. Therefore, inhibiting the truth bias may not be an influential factor in one's deception detection performance in such tasks.

Interestingly, it has been widely reported that autistic individuals at times struggle with inhibitory control (Geurts et al., 2014; Schmitt et al., 2018). Therefore, it is possible that this explains why, in comparison to non-autistic receivers, autistic individuals are worse at detecting deception in tasks where they have to make a truth-lie judgement about video, but are comparable in tasks where they have to identify the singular deceptive target from a set of stimuli.

While both sender demeanour and the need to inhibit receiver truth bias may go on to explain why autistic individuals performed relatively worse at deception detection on the Cheating Tasks and Box Lifts but not the TTL, it does not explain why a group difference was observed in the WP, which had a similar task design as TTL. It is however important to remember that while the TTL can be considered a valid deception detection task by the virtue of it correlating highly with the Cheating Task, which is a more established and widely used deception detection measure, similar evidence was not found for the WP. In fact, as speculated earlier, there is a possibility that other cognitive factors like imagination, generativity, and suspension of belief – all abilities that have been documented to be weak in autism (Craig & Baron-Cohen, 1999; Miller & Bugnariu, 2016; Samson & Hegenloh, 2010) – may have affected performance on the WP. Therefore, group differences observed in the WP may not purely be reflective of a difference in deception detection ability between autistic and non-autistic receivers, instead it may in part be a result of other cognitive difficulties autistic individuals have.

Therefore, it could be said that autistic individuals, while capable of detecting deception, show some difficulties *in comparison to* non-autistic individuals, but this varies based on the situation within which they are having to detect deception. But considering that in the real world factors affecting deception judgement cannot be controlled for, it is possible that autistic individuals may find themselves at a disadvantage still.

6.3.3. Contradictory evidence from supplementary findings

The findings from the additional analysis on the AQ-defined sample, however, must be taken into consideration before drawing conclusions. When those autistic individuals who scored below the cut-off of 26 and those non-autistic individuals who scored 32 or above on the AQ (recommended by Woodbury-Smith et al., 2005) were excluded, the difference in deception detection between groups found initially were no longer significant. In Chapter 4, this meant that autistic individuals were no longer worse at the Cheating task and the WP than non-autistic participants, and in Chapter 5 this meant that autistic receivers were distinguishing between Real and Pretend lifts as well as non-autistic receivers.

This could partly be because using the stringent exclusion criteria eliminated, or reduced to a great degree, the noise introduced by participants who were either inattentive, or did not identify with their diagnostic group as strongly, or were not actually autistic (in the case of excluded autistic participants). While there is no way to be sure of this, support from this comes from the fact that in Chapter 5, the AQ-defined autistic group showed stronger in-group identification and weaker out-group identification than they did in the original sample. This would suggest that the results regarding group differences found in the AQ-defined samples may be more valid than the full samples.

However, it cannot be ignored that by cleaning the data on the basis of the AQ, over a third in of the autistic sample in Chapter 4 and over a fourth of the autistic sample Chapter 5 was lost. Therefore, the group comparisons in both Chapter 4 and 5 suffered from a loss in power. Considering that the effect sizes were small to begin with in the original samples, the reduction in sample size could potentially be a reason as to why the differences between groups were not significant in the AQ-defined groups.

Therefore, the possibility that autistic individuals may be have some difficulties in deception detection when compared to non-autistic individuals cannot be rejected based solely on the supplementary findings. Considering that there is very little past literature on deception detection in autistic individuals, adults especially, it is therefore required that further investigation in this area is undertaken before concrete conclusions are drawn, especially considering the impact it can have on the wellbeing of autistic individuals.

6.4. What does this project tell us about the implications for autistic individuals?

Autistic individuals were tested on how well they detect deception in Chapter 4 and 5, and as discussed above, both chapters provided evidence that they are able to detect deception. However, in both instances, autistic receivers showed a reduced ability to detect deception when compared to non-autistic receivers in the original analyses. This suggests that although autistic individuals can detect deception to a certain extent, their skills are not as strong as in non-autistic individuals.

In chapter 4, the results revealed that deception detection ability predicted peer-victimization. As it can be tentatively inferred that autistic individuals are relatively weaker at deception detection than non-autistic individuals (Chapter 4 and 5), this becomes concerning as peer-victimization and instances of bullying targeting autistic individuals are already reported to be much higher than in their non-autistic counterparts (Sterzing et al., 2012). Moreover, deception detection further predicts psychological distress caused by peer-victimization, and mental health issues are much more common in autistic individuals (Lai et al., 2019). This may lead to the argument that training may have some merit, targeted at improving deception detection skills to attempt to increase quality of life in autistic individuals by potentially safeguarding autistic individuals from some bullying activity and helping alleviate some mental health issues caused due to that bullying.

In fact Ranick et al. (2013) demonstrated that deception detection training was possible for autistic children. They trained three autistic children to detect deceptive statements aimed to victimize them – either to exclude them from an activity, e.g., “Only people with brown hair can play this game.”, or as a means to take their possessions, e.g., “Your mom said I could have this.” (Ranick et al., 2013, p. 505). Training took multiple sessions and it was found that not only did autistic children learn to catch the lies, they continued doing so in subsequent sessions with novel lies told by their peers who were not part of the training, and maintained this ability one month after training as well. Van Tiel et al. (2021) too showed that autistic adults could learn to detect and produce deception in a game played against a computer that required little to no perspective taking.

However, van Tiel et al. (2021) concluded that although autistic individuals eventually learn to detect lies, they do so through explicit reasoning and conscious effort. This shows that autistic individuals can learn to detect deception through learning; however, this is likely through compensatory mechanisms. It is well documented that compensation increases psychological distress in autistic individuals (Bradley et al., 2021; J. Cook et al., 2021; Livingston et al., 2019). Additionally, it has been argued that interventions aimed at improving social skills are only "trying to promote a layer of compensation" (Livingston & Happé, 2019) and therefore might actually contribute to mental health issues. Thus, the question is if the benefits that come with interventions geared towards better deception detection – reduced peer victimization and mental health issues due to peer victimization – are worth the additional psychological distress and long-term negative consequences of compensatory learning of detection deception.

The other side of the argument is if there is an overemphasis on translating research outcomes into intervention for autistic individuals. Pellicano et al. (2014) identified that one of the areas that autistic individuals want research to advance in is to focus on changes outside of the autistic people themselves – adapting the environment to accommodate their needs and an exchange of knowledge so that there was better awareness about autism. Simply educating bullies about autism might not be a very effective way of reducing peer-victimization, but in order to improve autistic people's quality of lives in general, there is perhaps a need to *change the attitudes towards autistic individuals* rather than just focussing on how *autistic behaviours can be changed*. This need is highlighted by the unexpected finding of Chapter 5, where it was seen that autistic senders were generally rated to be more deceptive and this was not only true in situations where they were actively deceiving, but also when they were not being deceptive. While prior studies have showed that non-autistic individuals form a negative opinion about autistic individuals (Alkhaldi et al., 2021; Sasson et al., 2017), results here revealed that even autistic receivers were as likely to perceive autistic receivers as more deceptive regardless of their veracity as non-autistic receivers were. This shows that there is a general (mis)perception of autistic individuals which can be extremely detrimental, especially in high-stake situations.

Less than half the police force in Wales and England has had any training in dealing with autistic people (Crane et al., 2016). Combining this with the fact that autistic individuals are viewed as being more deceptive even when not deceiving, suggests that autistic individuals may be vulnerable to being wrongly accused or suspected of crimes, or not being believed when they are victims of crimes. Worryingly, from the findings of this project it appears highly likely that autistic individuals, already found to be victimized more (vs non-autistic individuals) partially due to possibly weaker deception detection skills, might not be considered truthful when they do report such instances. Therefore, this underlines the need for more awareness in the general public and sensitivity training for those who come in contact with autistic individuals as part of their job roles.

6.5. Limitations and future directions

This project was the first of its kind – investigating development of deception beyond childhood, development of deception in autistic individuals beyond childhood, and effects of neurotype-based group biases or the double empathy problem in deception detection. Nevertheless, the project did have some shortcomings, apart from the limitations that have been highlighted in relevant chapters, and the studies in this project leave room for improvement in future research.

6.5.1. Issues with tasks / measures used

In the empirical chapters, I have brought to attention the issues with some of the tasks used, such as the AQ and the AT-MCQ. Anecdotally, it has been brought to my attention that there are resources available online for autistic individuals about the popular tests that are administered in autism research and how to score well on these tasks (e.g. *Autism Tests | Embrace Autism*, n.d.). This emphasises the need to keep developing newer measures to test autistic individuals that are less widely known.

Considering that the performance on deception detection tasks such as the Cheating Task (T. R. Levine, 2007–2011) showed similar developmental trajectories to the mentalizing task used in Chapter 4, along with its sensitivity in picking up differences between autistic and non-autistic individuals as the mentalizing task does, deception detection tasks could potentially be used to record mentalizing ability. Deception (production) ability has been used as an indicator of mentalizing in children for decades (M. Chandler et al., 1989; Hala et al., 1991), therefore this is not a departure from past practices. It is relatively new (compared to established theory of mind tests) and therefore does not risk already being overused, and has the added benefit of being a better measure of mentalizing in real-world situations. This idea would require deception tasks like the Cheating Task to be further validated against different mentalizing tasks – both explicit and implicit – and investigations into other cognitive mechanisms that contribute to deception detection so that they can be controlled for, but the potential of developing deception detection tasks as measures of mentalizing ability is promising.

6.5.2. Issues with online testing

While optimistically the large sample size may aid in reducing noise in the data, the fact that it was recruited online may add to the concerns of generalizability. Rødgaard et al. (2022)

found that autistic individuals recruited online demonstrate high sampling bias, higher education levels and employment rates, and overall reduced generalisability of results drawn from such studies. Furthermore, online testing meant that there were no means to confirm the diagnosis of the participants. An attempt to mitigate the problem of participants who may self-diagnose was made by specifying a third category when asking for autism diagnosis apart from yes and no – "Yes - I self-identify as autistic but have not been diagnosed by a qualified clinician". It was assumed that any participants who were not diagnosed as autistic but identified as being autistic would hence select this option. Any participants who selected this option were screened out of the study either prior to taking part or their data was not included if they completed the study. Nonetheless, the deviating results when the participants were excluded if they did not meet autistic traits thresholds typical for their diagnostic groups, further adds to these concerns. Keeping these caveats in mind, the findings of this project require careful replication and any interpretations drawn must be done so with caution.

6.5.3. Inclusion of autistic senders and receivers

One limitation of the study that has been raised before is the lack of autistic senders in the novel deception production tasks. This had a knock-on effect throughout the project – as there were no autistic senders, the deception production study (Chapter 3) could only investigate cognitive underpinnings of deception production in a non-autistic sample. Although autistic traits in the general population show no discontinuity between autistic and non-autistic people (Happé & Frith, 2021), trait studies are not entirely suitable to draw concrete conclusions about cognition in autism (Sasson & Bottema-Beutel, 2022). Not having autistic senders also hampers the potential of the novel tasks being used as true double empathy tasks with senders of both neurotypes.

It is worthwhile to note however that the lack of autistic senders was not by choice. The CoViD-19 pandemic and associated restrictions affected participant recruitment and testing and made it impossible for videos from autistic senders to be collected in time for the studies in this project. Therefore, a logical future research direction would be to recruit autistic individuals for the lie production study, and thus also collect stimuli from autistic senders, and extend the studies and results in this project to draw more definite conclusion about deception in autism.

6.5.4. Deception is interactive, tasks were not

In the very beginning, I stressed how deception is an interactive activity and that studying it in vacuum is not representative. Therefore, it needs to be acknowledged that while some of the tasks were interactive from the senders' perspectives – senders in the novel tasks and the cheating task were actively deceiving another person – the tasks were not interactive from the receivers' perspectives. The independent receivers in Chapter 3 and the receivers in Chapter 4 and 5 all watched videos of senders that were recorded previously and they had no interaction with the senders. As mentioned in Chapter 2, receivers and senders feed off each other's cues in a deceptive episode, therefore the lack of interaction when detecting deception may lead to a slightly inaccurate measure of the receivers' deception detection skills.

That said, running an interactive experiment with a sample size to rival this project may not be feasible due to the costs and resources this would require. In addition, like all lab studies, the trade-off of lower ecological validity is greater experimental control. Where the deception detection tasks were low on interaction, it also meant that the stimuli were controlled for appropriateness, unlike past studies that have been more ecologically valid and have been plagued with low deception production rates and therefore inadequate stimuli for deception detection (Van Swol et al., 2012).

6.5.5. Alternative ways to record deception detection ability

With the novel tasks, this study attempted to record deception detection ability without using dichotomous responses of Lie / Truth, however that still left room for arbitrary guessing by the senders if they were unsure which of the three videos was the lie. This could have in turn inflated or suppressed their total scores and consequently introduced noise to the measure of their deception detection ability. In chapter 5, the box lifts recorded deception detection by using a scale instead: receivers rated how deceptive the stimuli were on a scale of 1 (real) to 5 (pretend). This possibly allowed receivers' judgement to incorporate how sure they were about their response by giving them an option to choose the degree to which they thought a stimulus was deceptive (or not). It is because of this that a more continuous measure of real-pretend or truth-lie than the traditional dichotomous truth / lie judgement or forced choice judgement like the one used in the WP and TTL might be a more sensitive way of recording people's deception detection ability. There is the possibility that receivers might resort to rating closer to the centre of the scale rather than the extremes, however, as observed in Chapter 5, receiver responses still significantly differentiate between real and pretend stimuli. Therefore, in the future, stimuli from the novel tasks and the Cheating Task could be adapted to be used alongside scale measures to reduce any noise in the data.

Future studies could also explore alternative ways to record deception detection ability by changing the time point at which the receiver makes the judgement. Cognition is dynamic and, in reality, stimuli are not presented in isolation and responses are not always made after the whole stimulus has been presented; instead, cognitive processing starts while stimuli are still being observed (Spivey & Dale, 2006). It can be advocated that likewise, day-to-day deception is not detected only after a sender has stopped speaking, but that receivers form opinions about a sender's honesty, truthfulness and believability throughout the duration of a deceptive episode. Therefore, it could be argued that standard judgments made by receivers after the fact might not capture the deception detection process holistically and that recording their response throughout the stimuli, i.e. synchronously, using a scale measure that allows participants to continuously change their responses on believability, honesty, truthfulness etc. might be better. Similar continuous response methods have been previously used in emotion research (e.g. Edgar, McRorie, Sneddon, 2012; McKeown & Sneddon, 2013). This also has the potential to allow for actual time course analysis of the process of deception detection and can possibly give an insight into similarities and dissimilarities in the process between certain groups, such as autistic and non-autistic individuals.

6.6. Closing statement

This thesis explored how deception production and deception detection are influenced by different characteristics of the person who is deceiving (sender) and the person who is being deceived (receiver). There is some evidence that deception judgement depends on senders, but that there are also individual differences in the receivers, such as mentalizing ability, autistic symptomatology, and age that predict deception judgement. Deception detection improves with advancement in mentalizing ability but only in the non-autistic group; autistic individuals are possibly weaker at deception detection than non-autistic individuals are and deception detection ability improves with age but again only in the non-autistic sample.

For the first time, the double empathy problem in the context of deception detection was explored and the investigations returned unexpected results: autistic individuals are judged as more deceptive regardless of their actual veracity or the receivers' neurotype. While this concurs with past literature on non-autistic people's negative judgement of autistic individuals, this has implications for how likely autistic individuals are to be believed in high stake situations such as in the criminal justice system.

This project is also the first time that the effect of deception detection on peer-victimization, and through peer-victimization its effect on mental health, has been empirically investigated. This gives rise to the possibility of new interventions involving deception detection training to protect against peer-victimization and associated mental health issues, especially for autistic individuals given their elevated rates of both. However, the potential benefits of such interventions must be weighed against the added psychological effects of long-term compensation that these interventions may introduce. Future studies should try to replicate the findings using enhanced methods, such as continuous synchronous deception ratings, in naïve samples, whose diagnosis can be verified more efficiently, in order to increase confidence in the results found here.

References

Aamodt, M., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner*, *15*, 6–11.

Abe, N. (2009). The neurobiology of deception: Evidence from neuroimaging and loss-of-function studies. *Current Opinion in Neurology*, *22*(6), 594–600. https://doi.org/10.1097/WCO.0b013e328332c3cf

Abe, N. (2011). How the Brain Shapes Deception: An Integrated Review of the Literature. *The Neuroscientist*, *17*(5), 560–574. https://doi.org/10.1177/1073858410393359

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. https://doi.org/10.1016/S0885-2014(00)00014-9

Aberson, C. L., & Howanski, L. M. (2002). Effects of self-esteem, status, and identificaton on two forms of intergroup bias. *Current Research in Social Psychology*, *7*(13), 225–242.

Abrahams, B. S., & Geschwind, D. H. (2008). Advances in autism genetics: On the threshold of a new neurobiology. *Nature Reviews Genetics*, *9*(5), Article 5. https://doi.org/10.1038/nrg2346

Adams, R. E., Fredstrom, B. K., Duncan, A. W., Holleb, L. J., & Bishop, S. L. (2014). Using Self- and Parent-Reports to Test the Association Between Peer Victimization and Internalizing Symptoms in Verbally Fluent Adolescents with ASD. *Journal of Autism and Developmental Disorders*, *44*(4), 861–872. https://doi.org/10.1007/s10803-013-1938-0

Adams, R., Rule, N., Franklin, R., Wang, E., Stevenson, M., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., & Ambady, N. (2009). Cross-cultural Reading the Mind in the

Eyes: An fMRI Investigation. *Journal of Cognitive Neuroscience*, *22*, 97–108. https://doi.org/10.1162/jocn.2009.21187

Ahmed, A. M. (2007). Group identity, social distance and intergroup bias. *Journal of Economic Psychology*, *28*(3), 324–337. https://doi.org/10.1016/j.joep.2007.01.007

Al Aseri, Z. A., Suriya, M. O., Hassan, H. A., Hasan, M., Sheikh, S. A., Al Tamimi, A., Alshathri, M., & Khalid, N. (2015). Reliability and validity of the Hospital Anxiety and Depression Scale in an emergency department in Saudi Arabia: A cross-sectional observational study. *BMC Emergency Medicine*, *15*, 28. https://doi.org/10.1186/s12873-015-0051-4

Alkhaldi, R. S., Sheppard, E., Burdett, E., & Mitchell, P. (2021). Do Neurotypical People Like or Dislike Autistic People? *Autism in Adulthood*, *3*(3), 275–279. https://doi.org/10.1089/aut.2020.0059

Allen, J., Bleiberg, E., Haslam-Hopwood, G., & Psyd. (2003). Mentalizing as a Compass for Treatment. *Bulletin of the Menninger Clinic, 67*.

Amaral, D. G. (Ed.). (2016). Reliable progress in autism research will depend on adequate research design, appropriately powered studies and efforts at collaboration to foster replication. In *Peril and Promise in Research on Autism: Reflections from the Editors of Autism Research* (p. 13). Wiley.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (5th edition). American Psychiatric Publishing.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Andersen, N. K., Rimvall, M. K., Jeppesen, P., Bentz, M., Jepsen, J. R. M., Clemmensen, L., Jacobsen, R. K., & Olsen, E. M. (2022). A psychometric investigation of the multiple-choice version of Animated Triangles Task to measure Theory of Mind in adolescence. *PLOS ONE*, *17*(3), e0264319. https://doi.org/10.1371/journal.pone.0264319

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Armitage, R. (2021). Bullying in children: Impact on child health. *BMJ Paediatrics Open*, *5*(1), e000939. https://doi.org/10.1136/bmjpo-2020-000939

Ashwood, K. L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F. S., Findon, J., Eklund, H., Spain, D., Wilson, C. E., Cadman, T., Young, S., Stoencheva, V., Murphy, C. M., Robertson, D., Charman, T., Bolton, P., Glaser, K., Asherson, P., … Murphy, D. G. (2016). Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological Medicine*, *46*(12), 2595–2604. https://doi.org/10.1017/S0033291716001082

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences*, *38*(2), 451–460. https://doi.org/10.1016/j.paid.2004.04.022

Autism tests | Embrace Autism. (n.d.). Retrieved 19 September 2022, from https://embrace-autism.com/autism-tests/

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*(1), 55–65. https://doi.org/10.1111/j.2044-8260.1993.tb01027.x

Bagnall, R., Russell, A., Brosnan, M., & Maras, K. (2022). Deceptive behaviour in autism: A scoping review. *Autism*, *26*(2), 293–307. https://doi.org/10.1177/13623613211057974

Bailey, A., Couteur, A. L., Gottesman, I., Bolton, P., 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. https://doi.org/10.1017/S0033291700028099

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Bal, E., Yerys, B. E., Sokoloff, J. L., Celano, M. J., Kenworthy, L., Giedd, J. N., & Wallace, G. L. (2013). Do social attribution skills improve with age in children with high functioning autism spectrum disorders? *Research in Autism Spectrum Disorders*, *7*(1), 9–16. https://doi.org/10.1016/j.rasd.2012.07.004

Barnicle, K. R. (2021). *The Role of Dispositional Factors on Veracity Judgments Accuracy and Bias* [M.S., Texas Woman's University]. https://www.proquest.com/docview/2600791280/abstract/755312067054361PQ/1

Baron-Cohen, S. (1992). Out of Sight or Out of Mind? Another Look at Deception in Autism. *Journal of Child Psychology and Psychiatry*, *33*(7), 1141–1155. https://doi.org/10.1111/j.1469-7610.1992.tb00934.x

Baron-Cohen, S. (2000a). Theory of mind and autism: A fifteen year review. In *Understanding other minds: Perspectives from developmental cognitive neuroscience, 2nd ed* (pp. 3–20). Oxford University Press.

Baron-Cohen, S. (2000b). Theory of mind and autism: A review. In *International Review of Research in Mental Retardation* (Vol. 23, pp. 169–184). Elsevier. https://doi.org/10.1016/S0074-7750(00)80010-5

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251. https://doi.org/10.1017/S0021963001006643

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/a:1005653411471

Battistelli, P., & Farneti, A. (2015). When the theory of mind would be very useful. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01449

Baum, S., Grazer, B., Moosekian, V., Sackheim, D., Baum, S., Nevins, D., Craft, E., Fain, S., Cary, A., Graziano, D., Ryan, S., Voll, D., & Maeda, S. (2009–2011). *Lie to Me*. Imagine Television; Samuel Baum Productions; MiddKid Productions.

BBFC | Ratings at a glance. (n.d.). Retrieved 26 September 2022, from https://www.bbfc.co.uk/rating/u

Bejerot, S., Eriksson, J. M., & Mörtberg, E. (2014). Social anxiety in adult autism spectrum disorder. *Psychiatry Research*, *220*(1), 705–707. https://doi.org/10.1016/j.psychres.2014.08.030

Bethlehem, J. (2010). Selection Bias in Web Surveys. *International Statistical Review*, *78*(2), 161–188. https://doi.org/10.1111/j.1751-5823.2010.00112.x

Betts, L. R., Houston, J. E., & Steer, O. L. (2015). Development of the Multidimensional Peer Victimization Scale–Revised (MPVS-R) and the Multidimensional Peer Bullying Scale (MPVS-RB). *The Journal of Genetic Psychology*, *176*(2), 93–109. https://doi.org/10.1080/00221325.2015.1007915

Bitsika, V., & Sharpley, C. F. (2018). The interaction of Matrix Reasoning and Social Motivation as predictors of Separation anxiety in boys with Autism Spectrum

Disorder. *International Journal of Developmental Neuroscience*, *67*(1), 6–13. https://doi.org/10.1016/j.ijdevneu.2018.03.006

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69–77. https://doi.org/10.1016/S0022-3999(01)00296-3

Blair, J. P., Levine, T. R., & Shaw, A. S. (2010). Content in Context Improves Deception Detection Accuracy. *Human Communication Research*, *36*(3), 423–442. https://doi.org/10.1111/j.1468-2958.2010.01382.x

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267–277. https://doi.org/10.1038/nrn2353

Blakemore, S.-J. (2012). Development of the social brain in adolescence. *Journal of the Royal Society of Medicine*, *105*(3), 111–116. https://doi.org/10.1258/jrsm.2011.110221

Blakemore, S.-J., Burnett, S., & Dahl, R. E. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping*, *31*(6), 926–933. https://doi.org/10.1002/hbm.21052

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), B25–B31. https://doi.org/10.1016/S0010-0277(00)00096-2

Bok, S. (1978). *Lying: Moral choice in public and private life*. Pantheon Books.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of Deception Judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, *134*(4), 477–492. https://doi.org/10.1037/0033-2909.134.4.477

Bond, G. D., Thompson, L. A., & Malloy, D. M. (2005). Vulnerability of Older Adults to Deception in Prison and Nonprison Contexts. *Psychology and Aging*, *20*(1), 60–70. https://doi.org/10.1037/0882-7974.20.1.60

Bottema-Beutel, K., Kapp, S. K., Lester, J. N., Sasson, N. J., & Hand, B. N. (2021). Avoiding Ableist Language: Suggestions for Autism Researchers. *Autism in Adulthood*, *3*(1), 18–29. https://doi.org/10.1089/aut.2020.0014

Bowerman, B. L., & O'Connell, R. T. (1990). *Linear Statistical Models: An Applied Approach* (2nd ed.). Duxbury.

Bowler, D. M. (1992). "Theory of Mind" in Asperger's Syndrome Dermot M. Bowler. *Journal of Child Psychology and Psychiatry*, *33*(5), 877–893. https://doi.org/10.1111/j.1469-7610.1992.tb01962.x

Bradley, L., Shaw, R., Baron-Cohen, S., & Cassidy, S. (2021). Autistic Adults' Experiences of Camouflaging and Its Perceived Impact on Mental Health. *Autism in Adulthood*, *3*(4), 320–329. https://doi.org/10.1089/aut.2020.0071

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980). The Truth-Deception Attribution: Effects of Familiarity on the Ability of Observers to Detect Deception. *Human Communication Research*, *6*(2), 99–110. https://doi.org/10.1111/j.1468-2958.1980.tb00130.x

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1982). Familiarity and lie detection: A replication and extension. *Western Journal of Speech Communication*, *46*(3), 276–290. https://doi.org/10.1080/10570318209374086

Brewer, N., Lucas, C. A., Lim, A., & Young, R. L. (2022). Detecting dodgy behaviour: The role of autism, autistic traits and theory of mind. *Autism*, 13623613221125564. https://doi.org/10.1177/13623613221125564

Brewer, N., Ying, A. B. W., Young, R. L., & Nah, Y.-H. (2018). Theory of mind and the detection of suspicious behavior. *Journal of Applied Research in Memory and Cognition*, *7*(1), 123–131. https://doi.org/10.1037/h0101817

Brewer, N., & Young, R. L. (2015). *Crime and Autism Spectrum Disorder: Myths and Mechanisms*. Jessica Kingsley Publishers.

Broadbent, J., Galic, I., & Stokes, M. A. (2013). Validation of Autism Spectrum Quotient Adult Version in an Australian Sample. *Autism Research and Treatment*, *2013,* e984205. https://doi.org/10.1155/2013/984205

Brown, C., Borduin, C., Dopp, A., & Mazurek, M. (2019). The social ecology of aggression in youths with autism spectrum disorder. *Autism Research*, *12*. https://doi.org/10.1002/aur.2157

Buescher, A. V. S., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of Autism Spectrum Disorders in the United Kingdom and the United States. *JAMA Pediatrics*, *168*(8), 721–728. https://doi.org/10.1001/jamapediatrics.2014.210

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal Deception Theory. *Communication Theory*, *6*(3), 203–242. https://doi.org/10.1111/j.1468-2885.1996.tb00127.x

Burgoon, J. K., Buller, D. B., & Floyd, K. (2001). Does Participation Affect Deception Success?: A Test of the Interactivity Principle. *Human Communication Research*, *27*(4), 503–534. https://doi.org/10.1111/j.1468-2958.2001.tb00791.x

Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S.-J. (2009). Development during Adolescence of the Neural Processing of Social Emotion. *Journal of Cognitive Neuroscience*, *21*(9), 1736–1750. https://doi.org/10.1162/jocn.2009.21121

Cappadocia, M. C., Weiss, J. A., & Pepler, D. (2012). Bullying Experiences Among Children and Youth with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(2), 266–277. https://doi.org/10.1007/s10803-011-1241-x

Carrión, R. E., Keenan, J. P., & Sebanz, N. (2010). A truth that's told with bad intent: An ERP study of deception. *Cognition*, *114*(1), 105–110. https://doi.org/10.1016/j.cognition.2009.05.014

Carter, S. (2009). Bullying of Students with Asperger Syndrome. *Issues in Comprehensive Pediatric Nursing*, *32*(3), 145–154. https://doi.org/10.1080/01460860903062782

Casartelli, L., Federici, A., Fumagalli, L., Cesareo, A., Nicoli, M., Ronconi, L., Vitale, A., Molteni, M., Rizzolatti, G., & Sinigaglia, C. (2020). Neurotypical individuals fail to understand action vitality form in children with autism spectrum disorder. *Proceedings of the National Academy of Sciences*, *117*(44), 27712–27718. https://doi.org/10.1073/pnas.2011311117

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*(8), 1839–1849. https://doi.org/10.1093/brain/awf189

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980–1008. https://doi.org/10.1037/a0035661

Chandler, J., Paolacci, G., Peer, E., Mueller, P., & Ratliff, K. A. (2015). Using Nonnaive Participants Can Reduce Effect Sizes. *Psychological Science*, *26*(7), 1131–1139. https://doi.org/10.1177/0956797615585115

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-Scale Deceit: Deception as a Marker of Two-, Three-, and Four-Year-Olds' Early Theories of Mind. *Child Development*, *60*(6), 1263–1277. https://doi.org/10.2307/1130919

Chang, F.-C., Lee, C.-M., Chiu, C.-H., Hsi, W.-Y., Huang, T.-F., & Pan, Y.-C. (2013). Relationships Among Cyberbullying, School Bullying, and Mental Health in Taiwanese Adolescents. *Journal of School Health*, *83*(6), 454–462. https://doi.org/10.1111/josh.12050

Chen, Y.-L., Senande, L. L., Thorsen, M., & Patten, K. (2021). Peer preferences and characteristics of same-group and cross-group social interactions among autistic and non-autistic adolescents. *Autism*, *25*(7), 1885–1900. https://doi.org/10.1177/13623613211005918

Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (2019). The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, *6*(10), 190232. https://doi.org/10.1098/rsos.190232

Christ, S. E., Kanne, S. M., & Reiersen, A. M. (2010). Executive function in individuals with subthreshold autism traits. *Neuropsychology*, *24*(5), 590–598. https://doi.org/10.1037/a0019176

Clare, D. D., & Levine, T. R. (2019). Documenting the Truth-Default: The Low Frequency of Spontaneous Unprompted Veracity Assessments in Deception Detection. *Human Communication Research*, *45*(3), 286–308. https://doi.org/10.1093/hcr/hqz001

Clementson, D. E. (2018). Truth Bias and Partisan Bias in Political Deception Detection. *Journal of Language and Social Psychology*, *37*(4), 407–430. https://doi.org/10.1177/0261927X17744004

Coggan, C., Bennett, S., Hooper, R., & Dickinson, P. (2003). Association between Bullying and Mental Health Status in New Zealand Adolescents. *International Journal of Mental Health Promotion*, *5*(1), 16–22. https://doi.org/10.1080/14623730.2003.9721892

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., Metzger, L. M., Shoushtari, C. S., Splinter, R., & Reich, W. (2003). Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*, *33*(4), 427–433. https://doi.org/10.1023/A:1025014929212

Constantino, J. N., & Todd, R. D. (2003). Autistic Traits in the General Population: A Twin Study. *Archives of General Psychiatry*, *60*(5), 524–530. https://doi.org/10.1001/archpsyc.60.5.524

Cook, J., Hull, L., Crane, L., & Mandy, W. (2021). Camouflaging in autism: A systematic review. *Clinical Psychology Review*, *89*, 102080. https://doi.org/10.1016/j.cpr.2021.102080

Cook, J. L., Blakemore, S.-J., & Press, C. (2013). Atypical basic movement kinematics in autism spectrum conditions. *Brain*, *136*(9), 2816–2824. https://doi.org/10.1093/brain/awt208

Crane, L., Maras, K. L., Hawken, T., Mulcahy, S., & Memon, A. (2016). Experiences of Autism Spectrum Disorder and Policing in England and Wales: Surveying Police and the Autism Community. *Journal of Autism and Developmental Disorders*, *46*(6), 2028–2041. https://doi.org/10.1007/s10803-016-2729-1

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism*, *24*(7), 1704–1712. https://doi.org/10.1177/1362361320919286

Crompton, C. J., Sharp, M., Axbey, H., Fletcher-Watson, S., Flynn, E. G., & Ropar, D. (2020). Neurotype-Matching, but Not Being Autistic, Influences Self and Observer

Ratings of Interpersonal Rapport. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.586171

Cunningham, A. B., & Schreibman, L. (2008). Stereotypy in autism: The importance of function. *Research in Autism Spectrum Disorders*, *2*(3), 469–479. https://doi.org/10.1016/j.rasd.2007.09.006

Curtis, D., Levine, T., Hart, C., & Serota, K. (2021). Does Information about the Frequency of Lying Impact Perceptions of Honesty? *North American Journal of Psychology*, *23*, 193–208.

Dasgupta, N. (2004). Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations. *Social Justice Research*, *17*(2), 143–169. https://doi.org/10.1023/B:SORE.0000027407.70241.15

Davidson, J., & Henderson, V. L. (2010). 'Coming out' on the spectrum: Autism, identity and disclosure. *Social & Cultural Geography*, *11*(2), 155–170. https://doi.org/10.1080/14649360903525240

Debey, E., Ridderinkhof, R. K., De Houwer, J., De Schryver, M., & Verschuere, B. (2015). Suppressing the truth as a mechanism of deception: Delta plots reveal the role of response inhibition in lying. *Consciousness and Cognition*, *37*, 148–159. https://doi.org/10.1016/j.concog.2015.09.005

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627

DePaulo, B. M. (2010). *The Hows and Whys of Lies*. CreateSpace Independent Publishing Platform.

DePaulo, B. M., Jordan, A., Irvine, A., & Laser, P. S. (1982). Age Changes in the Detection of Deception. *Child Development*, *53*(3), 701. https://doi.org/10.2307/1129383

DePaulo, B. M., & Kirkendol, S. E. (1989). The Motivational Impairment Effect in the Communication of Deception. In J. C. Yuille (Ed.), *Credibility Assessment* (pp. 51–70). Springer Netherlands. https://doi.org/10.1007/978-94-015-7856-1_3

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, *12*(3), 177–202. https://doi.org/10.1007/BF00987487

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

DePaulo, B. M., & Rosenthal, R. (1979). AGE CHANGES IN NONVERBAL DECODING SKILLS: EVIDENCE FOR INCREASING DIFFERENTIATION. *Merrill-Palmer Quarterly of Behavior and Development*, *25*(2), 145–150.

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of Mind, Gender, and Social Experiences in Middle Childhood. *Child Development*, *84*(3), 989–1003. https://doi.org/10.1111/cdev.12017

Diehl, M. (1990). The Minimal Group Paradigm: Theoretical Explanations and Empirical Findings. *European Review of Social Psychology*, *1*(1), 263–292. https://doi.org/10.1080/14792779108401864

Dillon, C. (2009). What's wrong. In *Highlights for Children*. Highlights for Children, Inc.

Dillon, C. (2010). That's silly. In *Highlights for Children*. Highlights for Children, Inc.

Ding, X. P., Heyman, G. D., Fu, G., Zhu, B., & Lee, K. (2018). Young children discover how to deceive in 10 days: A microgenetic study. *Developmental Science*, *21*(3), e12566. https://doi.org/10.1111/desc.12566

Ding, X. P., Sai, L., Fu, G., Liu, J., & Lee, K. (2014). Neural correlates of second-order verbal deception: A functional near-infrared spectroscopy (fNIRS) study. *NeuroImage*, *87*, 505–514. https://doi.org/10.1016/j.neuroimage.2013.10.023

Ding, X. P., Wellman, H. M., Wang, Y., Fu, G., & Lee, K. (2015). Theory-of-Mind Training Causes Honest Young Children to Lie. *Psychological Science*, *26*(11), 1812–1821. https://doi.org/10.1177/0956797615604628

Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology*, *38*(2), 205–211. https://doi.org/10.1111/1469-8986.3820205

Doherty-Sneddon, G., Whittle, L., & Riby, D. M. (2013). Gaze aversion during social style interactions in autism spectrum disorder and Williams syndrome. *Research in Developmental Disabilities*, *34*(1), 616–626. https://doi.org/10.1016/j.ridd.2012.09.022

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived Intragroup Variability as a Function of Group Status and Identification. *Journal of Experimental Social Psychology*, *31*(5), 410–436. https://doi.org/10.1006/jesp.1995.1018

Driskell, J. E. (2012). Effectiveness of deception detection training: A meta-analysis. *Psychology, Crime & Law*, *18*(8), 713–731. https://doi.org/10.1080/1068316X.2010.535820

Dubey, I., Ropar, D., & de C Hamilton, A. F. (2015). Measuring the value of social engagement in adults with and without autism. *Molecular Autism*, *6*(1), 35. https://doi.org/10.1186/s13229-015-0031-2

Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*(2), 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

Dunbar, N. E. (2017). Deception in Intergroup Communication. In N. E. Dunbar, *Oxford Research Encyclopedia of Communication*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.486

Dunbar, N. E., Gangi, K., Coveleski, S., Adams, A., Bernhold, Q., & Giles, H. (2016). When Is It Acceptable to Lie? Interpersonal and Intergroup Perspectives on Deception. *Communication Studies*, *67*(2), 129–146. https://doi.org/10.1080/10510974.2016.1146911

Dunbar, N. E., Jensen, M. L., Burgoon, J. K., Kelley, K. M., Harrison, K. J., Adame, B. J., & Bernard, D. R. (2015). Effects of Veracity, Modality, and Sanctioning on Credibility Assessment During Mediated and Unmediated Interviews. *Communication Research*, *42*(5), 649–674. https://doi.org/10.1177/0093650213480175

Dunham, Y., Srinivasan, M., Dotsch, R., & Barner, D. (2014). Religion insulates ingroup evaluations: The development of intergroup attitudes in India. *Developmental Science*, *17*(2), 311–319. https://doi.org/10.1111/desc.12105

Duran, G., Dochez, S., Tapiero, I., & Michael, G. A. (2020). Opinions, actions and emotions: Does the content of lies affect their detectability? *Psychology, Crime & Law*, *26*(10), 927–949. https://doi.org/10.1080/1068316X.2020.1742341

Eaves, L. C., & Ho, H. H. (2008). Young Adult Outcome of Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 739–747. https://doi.org/10.1007/s10803-007-0441-x

Edey, R., Cook, J., Brewer, R., Johnson, M. H., Bird, G., & Press, C. (2016). Interaction takes two: Typical adults exhibit mind-blindness towards those with autism spectrum disorder. *Journal of Abnormal Psychology*, *125*(7), 879–885. https://doi.org/10.1037/abn0000199

Ekman, P., & Friesen, W. V. (1969). Nonverbal Leakage and Clues to Deception. *Psychiatry*, *32*(1), 88–106.

Evans, A. D., & Lee, K. (2011). Verbal deception from late childhood to middle adolescence and its relation to executive functioning skills. *Developmental Psychology*, *47*(4), 1108–1116. https://doi.org/10.1037/a0023425

Fan, X., Griffin, D. J., & Tagg, E. P. (2022). Lie judgment trigger sensitivity and truth-bias: Truth default theory in intergroup communication. *Communication Quarterly*, *0*(0), 1–21. https://doi.org/10.1080/01463373.2022.2079994

Faso, D. J., Sasson, N. J., & Pinkham, A. E. (2015). Evaluating Posed and Evoked Facial Expressions of Emotion from Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *45*(1), 75–89. https://doi.org/10.1007/s10803-014-2194-7

Feeley, T. H., & deTurck, M. A. (1998). The Behavioral Correlates of Sanctioned and Unsanctioned Deceptive Communication. *Journal of Nonverbal Behavior*, *22*(3), 189–204. https://doi.org/10.1023/A:1022966505471

Feeley, T. H., deTurck, M. A., & Young, M. J. (1995). Baseline familiarity in lie detection. *Communication Research Reports*, *12*(2), 160–169. https://doi.org/10.1080/08824099509362052

Feldman, R. (2009). *The Liar in Your Life: How Lies Work and What They Tell Us About Ourselves*. Random House.

Feldman, R. S., Tomasian, J. C., & Coats, E. J. (1999). Nonverbal Deception Abilities and Adolescents' Social Competence: Adolescents with Higher Social Skills are Better Liars. *Journal of Nonverbal Behavior*, *23*(3), 237–249. https://doi.org/10.1023/A:1021369327584

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, *65*(6), 591–598. https://doi.org/10.1203/PDR.0b013e31819e7203

Ford, R., King, T., Priest, N., & Kavanagh, A. (2017). Bullying and mental health and suicidal behaviour among 14- to 15-year-olds in a representative sample of Australian children. *Australian & New Zealand Journal of Psychiatry*, *51*(9), 897–908. https://doi.org/10.1177/0004867417700275

Frank, M. G. (2008). Research Methods in Detecting Deception Research. In J. Harrigan, R. Rosenthal, & K. Scherer (Eds.), *The New Handbook of Methods in Nonverbal Behavior Research* (pp. 341–368). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198529620.003.0009

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, *72*(6), 1429–1439. https://doi.org/10.1037/0022-3514.72.6.1429

Frank, M. G., & Svetieva, E. (2015). Microexpressions and Deception. In M. K. Mandal & A. Awasthi (Eds.), *Understanding Facial Expressions in Communication: Cross-cultural and Multidisciplinary Perspectives* (pp. 227–242). Springer India. https://doi.org/10.1007/978-81-322-1934-7_11

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, U. (2004). Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *45*(4), 672–686. https://doi.org/10.1111/j.1469-7610.2004.00262.x

Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, *50*(1), 115–132. https://doi.org/10.1016/0010-0277(94)90024-8

Frith, U., & Happé, F. (1999). Theory of Mind and Self-Consciousness: What Is It Like to Be Autistic? *Mind & Language*, *14*(1), 82–89. https://doi.org/10.1111/1468-0017.00100

Fuller, C. M., Biros, D. P., Twitchell, D. P., & Wilson, R. L. (2015). Real-World Deception and the Impact of Severity. *Journal of Computer Information Systems*, *55*(2), 59–67. https://doi.org/10.1080/08874417.2015.11645757

Gagné, N., & Franzen, L. (2021). *How to Run Behavioural Experiments Online: Best Practice Suggestions for Cognitive Psychology and Neuroscience*. PsyArXiv. https://doi.org/10.31234/osf.io/nt67j

Ganis, G., & Keenan, J. P. (2009). The cognitive neuroscience of deception. *Social Neuroscience*, *4*(6), 465–472. https://doi.org/10.1080/17470910802507660

Gee, J., & Button, M. (2021). *The financial cost of fraud 2021* (p. 30). University of Plymouth,. https://f.datasrvr.com/fr1/521/90994/0031_Financial_Cost_of_Fraud_2021_v5.pdf

Geelhand, P., Papastamou, F., Deliens, G., & Kissine, M. (2021). Judgments of spoken discourse and impression formation of neurotypical and autistic adults. *Research in Autism Spectrum Disorders*, *82*, 101742. https://doi.org/10.1016/j.rasd.2021.101742

George, J. F., Tilley, P., & Giordano, G. (2014). Sender credibility and deception detection. *Computers in Human Behavior*, *35*, 1–11. https://doi.org/10.1016/j.chb.2014.02.027

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. https://doi.org/10.1002/aur.1369

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*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00438

Glenwright, M., Scott, R. M., Bilevicius, E., Pronovost, M., & Hanlon-Dearman, A. (2021). Children With Autism Spectrum Disorder Can Attribute False Beliefs in a Spontaneous-Response Preferential-Looking Task. *Frontiers in Communication*, *6*. https://www.frontiersin.org/article/10.3389/fcomm.2021.669985

Global Deception Research Team. (2006). A World of Lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74. https://doi.org/10.1177/0022022105282295

Gombos, V. A. (2006). The Cognition of Deception: The Role of Executive Processes in Producing Lies. *Genetic, Social, and General Psychology Monographs*, *132*(3), 197–214. https://doi.org/10.3200/MONO.132.3.197-214

Gorilla Support (Director). (2022, May 5). *Gorilla Presents...Improving Participant Engagement*. https://www.youtube.com/watch?v=s95XIGkx_LM

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, *37*(4), 613–627. https://doi.org/10.1007/s10803-006-0280-1

Gowen, E., Stanley, J., & Miall, R. C. (2008). Movement interference in autism-spectrum disorder. *Neuropsychologia*, *46*(4), 1060–1068. https://doi.org/10.1016/j.neuropsychologia.2007.11.004

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, *52*, 15–23. https://doi.org/10.1016/j.jesp.2013.12.001

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling Theories of Mind: Deception as a Marker of Three-Year-Olds' Understanding of False Belief. *Child Development*, *62*(1), 83–97. https://doi.org/10.2307/1130706

Hamilton, A. F. de C., & Kumano, S. (Directors). (personal communication). *Box Lifts Video Set*.

Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., Dolan, R., Frackowiak, R., & Frith, C. (1996). 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. *NeuroReport*, *8*(1), 197–201.

Happé, F., & Frith, U. (2021). Dimensional or Categorical Approaches to Autism? Both are Needed. A Reply to Nick Chown and Julia Leatherland. *Journal of Autism and Developmental Disorders*, *51*(2), 752–753. https://doi.org/10.1007/s10803-020-04728-5

Hartwig, M., & Bond Jr., C. F. (2014). Lie Detection from Multiple Cues: A Meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661–676. https://doi.org/10.1002/acp.3052

Hauser, D., Paolacci, G., & Chandler, J. (2019). Common Concerns with MTurk as a Participant Pool: Evidence and Solutions. In *Handbook of Research Methods in Consumer Psychology*. Routledge.

Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (3rd edition). Guilford Press.

Heasman, B., & Gillespie, A. (2019). Neurodivergent intersubjectivity: Distinctive features of how autistic people create shared understanding. *Autism*, *23*(4), 910–921. https://doi.org/10.1177/1362361318785172

Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia*, *47*(4), 1004–1012. https://doi.org/10.1016/j.neuropsychologia.2008.10.023

Herrmann, C. (1997). International experiences with the Hospital Anxiety and Depression Scale-A review of validation data and clinical results. *Journal of Psychosomatic Research*, *42*(1), 17–41. https://doi.org/10.1016/S0022-3999(96)00216-4

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup Bias. *Annual Review of Psychology*, *53*(1), 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Holden, R., Mueller, J., McGowan, J., Sanyal, J., Kikoler, M., Simonoff, E., Velupillai, S., & Downs, J. (2020). Investigating Bullying as a Predictor of Suicidality in a Clinical Sample of Adolescents with Autism Spectrum Disorder. *Autism Research*, *13*(6), 988–997. https://doi.org/10.1002/aur.2292

Holt, M. K., Kaufman Kantor, G., & Finkelhor, D. (2008). Parent/Child Concordance about Bullying Involvement and Family Characteristics Related to Bullying and Peer Victimization. *Journal of School Violence*, *8*(1), 42–63. https://doi.org/10.1080/15388220802067813

Hong, Y., & Ratner, K. G. (2021). Minimal but not meaningless: Seemingly arbitrary category labels can imply more than group membership. *Journal of Personality and Social Psychology*, *120*(3), 576–600. https://doi.org/10.1037/pspa0000255

Hubbard, D. J., Faso, D. J., Assmann, P. F., & Sasson, N. J. (2017). Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Research*, *10*(12), 1991–2001. https://doi.org/10.1002/aur.1847

Huepe, D., Roca, M., Salas, N., Canales-Johnson, A., Rivera-Rei, Á. A., Zamorano, L., Concepción, A., Manes, F., & Ibañez, A. (2011). Fluid Intelligence and Psychosocial

Outcome: From Logical Problem Solving to Social Adaptation. *PLoS ONE*, *6*(9), e24858. https://doi.org/10.1371/journal.pone.0024858

Humphrey, N., & Symes, W. (2010). Responses to bullying and use of social support among pupils with autism spectrum disorders (ASDs) in mainstream schools: A qualitative study. *Journal of Research in Special Educational Needs*, *10*(2), 82–90. https://doi.org/10.1111/j.1471-3802.2010.01146.x

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., & Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the Autism Spectrum Quotient (AQ) in a non-clinical sample. *Personality and Individual Differences*, *43*(7), 1938–1949. https://doi.org/10.1016/j.paid.2007.06.012

Hwang, S., Kim, Y. S., Koh, Y.-J., & Leventhal, B. L. (2018). Autism Spectrum Disorder and School Bullying: Who is the Victim? Who is the Perpetrator? *Journal of Autism and Developmental Disorders*, *48*(1), 225–238. https://doi.org/10.1007/s10803-017-3285-z

Im-Bolter, N., Agostino, A., & Owens-Jaffray, K. (2016). Theory of mind in middle childhood and early adolescence: Different from before? *Journal of Experimental Child Psychology*, *149*, 98–115. https://doi.org/10.1016/j.jecp.2015.12.006

Ingersoll, B., Hopwood, C. J., Wainer, A., & Brent Donnellan, M. (2011). A Comparison of Three Self-Report Measures of the Broader Autism Phenotype in a Non-Clinical Sample. *Journal of Autism and Developmental Disorders*, *41*(12), 1646–1657. https://doi.org/10.1007/s10803-011-1192-2

Jaarsma, P., Gelhaus, P., & Welin, S. (2012). Living the categorical imperative: Autistic perspectives on lying and truth telling–between Kant and care ethics. *Medicine, Health Care and Philosophy*, *15*(3), 271–277. https://doi.org/10.1007/s11019-011-9363-7

James, E. (2019). *Pre-experiment Sound Check* [Gorilla Experiment Builder]. University of York. https://app.gorilla.sc/openmaterials/42292

Jansiewicz, E. M., Goldberg, M. C., Newschaffer, C. J., Denckla, M. B., Landa, R., & Mostofsky, S. H. (2006). Motor Signs Distinguish Children with High Functioning Autism and Asperger's Syndrome from Controls. *Journal of Autism and Developmental Disorders*, *36*(5), 613–621. https://doi.org/10.1007/s10803-006-0109-y

Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2004). The Right to Do Wrong: Lying to Parents Among Adolescents and Emerging Adults. *Journal of Youth and Adolescence*, *33*(2), 101–112. https://doi.org/10.1023/B:JOYO.0000013422.48100.5a

Jones, C. R. G., Swettenham, J., Charman, T., Marsden, A. J. S., Tregay, J., Baird, G., Simonoff, E., & Happé, F. (2011). No evidence for a fundamental visual motion processing deficit in adolescents with autism spectrum disorders. *Autism Research*, *4*(5), 347–357. https://doi.org/10.1002/aur.209

Joseph, R. M., & Tager–Flusberg, H. (2004). The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology*, *16*(1), 137–155. https://doi.org/10.1017/S095457940404444X

Joseph, S., & Stockton, H. (2018). The multidimensional peer victimization scale: A systematic review. *Aggression and Violent Behavior*, *42*, 96–114. https://doi.org/10.1016/j.avb.2018.07.009

Kassin, S. M., & Fong, C. T. (1999). 'I'm innocent!': Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, *23*(5), 499–516. https://doi.org/10.1023/A:1022330011811

Keating, C. F., & Heltman, K. R. (1994). Dominance and Deception in Children and Adults: Are Leaders the Best Misleaders? *Personality and Social Psychology Bulletin*, *20*(3), 312–321. https://doi.org/10.1177/0146167294203009

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Kireev, M., Korotkov, A., Medvedeva, N., Masharipov, R., & Medvedev, S. (2017). Deceptive but Not Honest Manipulative Actions Are Associated with Increased Interaction between Middle and Inferior Frontal gyri. *Frontiers in Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fnins.2017.00482

Klindt, D., Devaine, M., & Daunizeau, J. (2017). Does the way we read others' mind change over the lifespan? Insights from a massive web poll of cognitive skills from childhood to late adulthood. *Cortex, 86*, 205–215. https://doi.org/10.1016/j.cortex.2016.09.009

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling, Fourth Edition*. Guilford Publications.

Koelkebeck, K., Liedtke, C., Kohl, W., Alferink, J., & Kret, M. E. (2017). Attachment style moderates theory of mind abilities in depression. *Journal of Affective Disorders*, *213*, 156–160. https://doi.org/10.1016/j.jad.2017.02.021

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive Development*, *46*, 97–111. https://doi.org/10.1016/j.cogdev.2017.09.001

Kumpulainen, K., Räsänen, E., & Puura, K. (2001). Psychiatric disorders and the use of mental health services among children involved in bullying. *Aggressive Behavior, 27*(2), 102–110. https://doi.org/10.1002/ab.3

Kunihira, Y., Senju, A., Dairoku, H., Wakabayashi, A., & Hasegawa, T. (2006). 'Autistic' Traits in Non-Autistic Japanese Populations: Relationships with Personality Traits and Cognitive Ability. *Journal of Autism and Developmental Disorders*, *36*(4), 553–566. https://doi.org/10.1007/s10803-006-0094-1

Kuusikko, S., Pollock-Wurman, R., Jussila, K., Carter, A. S., Mattila, M.-L., Ebeling, H., Pauls, D. L., & Moilanen, I. (2008). Social Anxiety in High-functioning Children and Adolescents with Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *38*(9), 1697–1709. https://doi.org/10.1007/s10803-008-0555-9

Lai, M.-C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry*, *6*(10), 819–829. https://doi.org/10.1016/S2215-0366(19)30289-5

Laycock, R. (2018, November 6). *How many of us lie on insurance applications?* Finder.Com. https://www.finder.com/lying-on-insurance

Le, H. T. H., Nguyen, H. T., Campbell, M. A., Gatton, M. L., Tran, N. T., & Dunne, M. P. (2017). Longitudinal associations between bullying and mental health among adolescents in Vietnam. *International Journal of Public Health*, *62*(1), 51–61. https://doi.org/10.1007/s00038-016-0915-8

Leduc, K., Williams, S., Gomez-Garibello, C., & Talwar, V. (2017). The contributions of mental state understanding and executive functioning to preschool-aged children's lie-telling. *British Journal of Developmental Psychology*, *35*(2), 288–302. https://doi.org/10.1111/bjdp.12163

Leekam, S. R., & Prior, M. (1994). Can autistic children distinguish lies from jokes? A second look at second-order belief attribution. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *35*(5), 901–915. https://doi.org/10.1111/j.1469-7610.1994.tb02301.x

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*(2), 152–168. https://doi.org/10.1002/ejsp.2079

Leng, H., Wang, Y., Li, Q., Yang, L., & Sun, Y. (2019). Sophisticated Deception in Junior Middle School Students: An ERP Study. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02675

Levine, T., Clare, D., Greene, T., Serota, K., & Park, H. (2014). The effects of truth-lie base-rate on interactive deception detection accuracy: The Park-Levine probability model versus interpersonal deception theory. *Human Communication Research*.

Levine, T. R. (2007–2011). *NSF funded cheating tape interviews*. East Lansing.

Levine, T. R. (2010). A Few Transparent Liars Explaining 54% Accuracy in Deception Detection Experiments. *Annals of the International Communication Association*, *34*(1), 41–61. https://doi.org/10.1080/23808985.2010.11679095

Levine, T. R. (2014). Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. https://doi.org/10.1177/0261927X14535916

Levine, T. R. (2016). Examining Sender and Judge Variability in Honesty Assessments and Deception Detection Accuracy: Evidence for a Transparent Liar but No Evidence of Deception-General Ability. *Communication Research Reports*, *33*(3), 188–194. https://doi.org/10.1080/08824096.2016.1186629

Levine, T. R. (2018). Ecological Validity and Deception Detection Research Design. *Communication Methods and Measures*, *12*(1), 45–54. https://doi.org/10.1080/19312458.2017.1411471

Levine, T. R., Clare, D. D., Blair, J. P., McCornack, S., Morrison, K., & Park, H. S. (2014). Expertise in Deception Detection Involves Actively Prompting Diagnostic

Information Rather Than Passive Behavioral Observation. *Human Communication Research*, *40*(4), 442–462. https://doi.org/10.1111/hcre.12032

Levine, T. R., Feeley, T. H., McCornack, S. A., Hughes, M., & Harms, C. M. (2005). Testing the Effects of Nonverbal Behavior Training on Accuracy in Deception Detection with the Inclusion of a Bogus Training Control Group. *Western Journal of Communication*, *69*(3), 203–217. https://doi.org/10.1080/10570310500202355

Levine, T. R., Kim, R. K., & Hamel, L. M. (2010). People Lie for a Reason: Three Experiments Documenting the Principle of Veracity. *Communication Research Reports*, *27*(4), 271–285. https://doi.org/10.1080/08824096.2010.496334

Levine, T. R., & Knapp, M. L. (2018). Lying and Deception in Close Relationships. In A. L. Vangelisti & D. Perlman (Eds.), *The Cambridge Handbook of Personal Relationships* (2nd ed., pp. 329–340). Cambridge University Press. https://doi.org/10.1017/9781316417867.026

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the “veracity effect”. *Communication Monographs*, *66*(2), 125–144. https://doi.org/10.1080/03637759909376468

Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., Shim, J. C., & Lee, J. H. (2011). Sender Demeanor: Individual Differences in Sender Believability Have a Powerful Impact on Deception Detection Judgments. *Human Communication Research*, *37*(3), 377–403. https://doi.org/10.1111/j.1468-2958.2011.01407.x

Levine, T. R., Shaw, A., & Shulman, H. C. (2010). Increasing Deception Detection Accuracy with Strategic Questioning. *Human Communication Research*, *36*(2), 216–231. https://doi.org/10.1111/j.1468-2958.2010.01374.x

Levine, T., Shaw, A., & Shulman, H. C. (2010). Increasing Deception Detection Accuracy with Strategic Questioning. *Human Communication Research*, *36*(2), 216–231. https://doi.org/10.1111/j.1468-2958.2010.01374.x

Li, A. S., Kelley, E. A., Evans, A. D., & Lee, K. (2011). Exploring the Ability to Deceive in Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *41*(2), 185–195. https://doi.org/10.1007/s10803-010-1045-4

Li, L., Chen, X., & Li, H. (2020). Bullying victimization, school belonging, academic engagement and achievement in adolescents in rural China: A serial mediation model. *Children and Youth Services Review*, *113*, 104946. https://doi.org/10.1016/j.childyouth.2020.104946

Libster, N., Knox, A., Engin, S., Geschwind, D., Parish-Morris, J., & Kasari, C. (2022). Personal victimization experiences of autistic and non-autistic children. *Molecular Autism*, *13*(1), 51. https://doi.org/10.1186/s13229-022-00531-4

Lim, A., Young, R. L., & Brewer, N. (2022). Autistic Adults May Be Erroneously Perceived as Deceptive and Lacking Credibility. *Journal of Autism and Developmental Disorders*, *52*(2), 490–507. https://doi.org/10.1007/s10803-021-04963-4

Lisofsky, N., Kazzer, P., Heekeren, H. R., & Prehn, K. (2014). Investigating socio-cognitive processes in deception: A quantitative meta-analysis of neuroimaging studies. *Neuropsychologia*, *61*, 113–122. https://doi.org/10.1016/j.neuropsychologia.2014.06.001

Little, L. (2002). Middle-Class Mothers' Perceptions of Peer and Sibling Victimization Among Children with Asperger's Syndrome and Nonverbal Learning Disorders. *Issues in Comprehensive Pediatric Nursing*, *25*(1), 43–57. https://doi.org/10.1080/014608602753504847

Littlepage, G. E., Tang, D. W., & Pineault, M. A. (1986). Nonverbal and Content Factors in the Detection of Deception in Planned and Spontaneous Communications. *Journal of Social Behavior and Personality*, *1*(3), 439–450.

Liu, M. C.-C., Lan, C.-C., Hsu, J.-W., Huang, K.-L., & Chen, Y.-S. (2014). Bullying victimization and conduct problems among high school students in Taiwan: Focus on fluid intelligence, mood symptoms and associated psychosocial adjustment. *Children and Youth Services Review*, *47*(P3), 231–238.

Livingston, L. A., Colvert, E., Team, the S. R. S., Bolton, P., & Happé, F. (2019). Good social skills despite poor theory of mind: Exploring compensation in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *60*(1), 102–110. https://doi.org/10.1111/jcpp.12886

Livingston, L. A., Shah, P., White, S. J., & Happé, F. (2021). Further developing the Frith–Happé animations: A quicker, more objective, and web-based test of theory of mind for autistic and neurotypical adults. *Autism Research*, *14*(9), 1905–1912. https://doi.org/10.1002/aur.2575

Lloyd, E. P., Hugenberg, K., McConnell, A. R., Kunstman, J. W., & Deska, J. C. (2017). Black and White Lies: Race-Based Biases in Deception Judgments. *Psychological Science*, *28*(8), 1125–1136. https://doi.org/10.1177/0956797617705399

Locke, J., Ishijima, E. H., Kasari, C., & London, N. (2010). Loneliness, friendship quality and the social networks of adolescents with high-functioning autism in an inclusive school setting. *Journal of Research in Special Educational Needs*, *10*(2), 74–81. https://doi.org/10.1111/j.1471-3802.2010.01148.x

Luke, T. J. (2019). Lessons From Pinocchio: Cues to Deception May Be Highly Exaggerated. *Perspectives on Psychological Science*, *14*(4), 646–671. https://doi.org/10.1177/1745691619838258

Lundqvist, L.-O., & Lindner, H. (2017). Is the Autism-Spectrum Quotient a Valid Measure of Traits Associated with the Autism Spectrum? A Rasch Validation in Adults with and Without Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *47*(7), 2080–2091. https://doi.org/10.1007/s10803-017-3128-y

Ma, F., Evans, A. D., Liu, Y., Luo, X., & Xu, F. (2015). To lie or not to lie? The influence of parenting and theory-of-mind understanding on three-year-old children's honesty. *Journal of Moral Education*, *44*(2), 198–212. https://doi.org/10.1080/03057240.2015.1023182

Maboyana, Y., & Sekaja, L. (2015). Exploring the bullying behaviours experienced by South African domestic workers. *Journal of Psychology in Africa*, *25*(2), 114–120. https://doi.org/10.1080/14330237.2015.1021511

Madipakkam, A. R., Rothkirch, M., Dziobek, I., & Sterzer, P. (2017). Unconscious avoidance of eye contact in autism spectrum disorder. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-13945-5

Maes, J. H., Vissers, C. T., Egger, J. I., & Eling, P. A. (2013). On the relationship between autistic traits and executive functioning in a non-clinical Dutch student population. *Autism*, *17*(4), 379–389. https://doi.org/10.1177/1362361312442009

Mann, S., Vrij, A., & Bull, R. (2004). Detecting True Lies: Police Officers' Ability to Detect Suspects' Lies. *The Journal of Applied Psychology*, *89*, 137–149. https://doi.org/10.1037/0021-9010.89.1.137

Maras, K., Marshall, I., & Sands, C. (2019). Mock Juror Perceptions of Credibility and Culpability in an Autistic Defendant. *Journal of Autism and Developmental Disorders*, *49*(3), 996–1010. https://doi.org/10.1007/s10803-018-3803-7

Mathersul, D., McDonald, S., & Rushby, J. A. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal* *of Clinical and Experimental Neuropsychology*, *35*(6), 655–668. https://doi.org/10.1080/13803395.2013.809700

McCornack, S. A. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In *Message production: Advances in communication theory* (pp. 91–126). Lawrence Erlbaum Associates Publishers.

McCornack, S. A., & Parks, M. R. (1990). What Women Know that Men don't: Sex Differences in Determining the Truth Behind Deceptive Messages. *Journal of Social and Personal Relationships*, *7*(1), 107–118. https://doi.org/10.1177/0265407590071006

McLoughlin, N., & Over, H. (2017). Young Children Are More Likely to Spontaneously Attribute Mental States to Members of Their Own Group. *Psychological Science*, *28*(10), 1503–1509. https://doi.org/10.1177/0956797617710724

Meissner, C. A., & Kassin, S. M. (2002). 'He's guilty!': Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*(5), 469–480. https://doi.org/10.1023/A:1020278620751

Menard, S. (2002). *Applied Logistic Regression Analysis*. SAGE.

Meyer, P. (2010). *Liespotting: Proven Techniques to Detect Deception*. St. Martin's Publishing Group.

Meyer, P. (Director). (2011, July). *How to spot a liar*. https://www.ted.com/talks/pamela_meyer_how_to_spot_a_liar

Michopoulos, I., Douzenis, A., Kalkavoura, C., Christodoulou, C., Michalopoulou, P., Kalemi, G., Fineti, K., Patapis, P., Protopapas, K., & Lykouras, L. (2008). Hospital Anxiety and Depression Scale (HADS): Validation in a Greek general hospital sample. *Annals of General Psychiatry*, *7*(1), 4. https://doi.org/10.1186/1744-859X-7-4

Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication* (pp. xi, 131). Sage Publications, Inc.

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773. https://doi.org/10.1037/a0016854

Milton, D. (2018). *The double empathy problem*. Network Autism, National Autistic Society. https://www.autism.org.uk/advice-and-guidance/professional-practice/double-empathy

Milton, D. E. M. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society*, *27*(6), 883–887. https://doi.org/10.1080/09687599.2012.710008

Mintah, K., & Parlow, S. E. (2018). Are you flirting with me? Autistic traits, theory of mind, and inappropriate courtship. *Personality and Individual Differences*, *128*, 100–106. https://doi.org/10.1016/j.paid.2018.02.028

Mitchell, P., Sheppard, E., & Cassidy, S. (2021). Autism and the double empathy problem: Implications for development and mental health. *British Journal of Developmental Psychology*, *39*(1), 1–18. https://doi.org/10.1111/bjdp.12350

Moessnang, C., Baumeister, S., Tillmann, J., Goyard, D., Charman, T., Ambrosino, S., Baron-Cohen, S., Beckmann, C., Bölte, S., Bours, C., Crawley, D., Dell'Acqua, F., Durston, S., Ecker, C., Frouin, V., Hayward, H., Holt, R., Johnson, M., Jones, E., … Meyer-Lindenberg, A. (2020). Social brain activation during mentalizing in a large autism cohort: The Longitudinal European Autism Project. *Molecular Autism*, *11*(1), 17. https://doi.org/10.1186/s13229-020-0317-x

Molenberghs, P., & Louis, W. R. (2018). Insights From fMRI Studies Into Ingroup Bias. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01868

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences*, *108*(7), 2688–2692. https://doi.org/10.1073/pnas.1011734108

Morgan, C. J., LeSage, J. B., & Kosslyn, S. M. (2009). Types of deception revealed by individual differences in cognitive abilities. *Social Neuroscience*, *4*(6), 554–569. https://doi.org/10.1080/17470910802299987

Möricke, E., Buitelaar, J. K., & Rommelse, N. N. J. (2016). Do We Need Multiple Informants When Assessing Autistic Traits? The Degree of Report Bias on Offspring, Self, and Spouse Ratings. *Journal of Autism and Developmental Disorders*, *46*(1), 164–175. https://doi.org/10.1007/s10803-015-2562-y

Moriguchi, Y., Ohnishi, T., Mori, T., Matsuda, H., & Komaki, G. (2007). Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. *Psychiatry and Clinical Neurosciences*, *61*(4), 355–363. https://doi.org/10.1111/j.1440-1819.2007.01687.x

Morrison, K. E., DeBrabander, K. M., Faso, D. J., & Sasson, N. J. (2019). Variability in first impressions of autistic adults made by neurotypical raters is driven more by characteristics of the rater than by characteristics of autistic adults. *Autism*, *23*(7), 1817–1829. https://doi.org/10.1177/1362361318824104

Morrison, K. E., DeBrabander, K. M., Jones, D. R., Faso, D. J., Ackerman, R. A., & Sasson, N. J. (2020). Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism*, *24*(5), 1067–1080. https://doi.org/10.1177/1362361319892701

Mottron, L., & Bzdok, D. (2020). Autism spectrum heterogeneity: Fact or artifact? *Molecular Psychiatry*, *25*(12), Article 12. https://doi.org/10.1038/s41380-020-0748-y

Munro, M. J., & Derwing, T. M. (1995). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., Hammond, N., Murphy, D., & Happé, F. (2017). A new test of advanced theory of mind: The "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders. *Autism Research*, *10*(6), 1120–1132. https://doi.org/10.1002/aur.1744

Muszbek, K., Szekely, A., Balogh, É. M., Molnár, M., Rohánszky, M., Ruzsa, Á., Varga, K., Szöllosi, M., & Vadász, P. (2006). Validation of the Hungarian Translation of Hospital Anxiety and Depression Scale. *Quality of Life Research*, *15*(4), 761–766. https://doi.org/10.1007/s11136-005-3627-8

Nancarrow, A. F., Gilpin, A. T., Thibodeau, R. B., & Farrell, C. B. (2018). Knowing what others know: Linking deception detection, emotion knowledge, and Theory of Mind in preschool. *Infant and Child Development*, *27*(5), e2097. https://doi.org/10.1002/icd.2097

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying Behaviors Among US Youth: Prevalence and Association With Psychosocial Adjustment. *JAMA*, *285*(16), 2094. https://doi.org/10.1001/jama.285.16.2094

Nelson, L., Crawford, H., Reid, D., Moss, J., & Oliver, C. (2017). An experimental study of executive function and social impairment in Cornelia de Lange syndrome. *Journal of Neurodevelopmental Disorders*, *9*, 33. https://doi.org/10.1186/s11689-017-9213-x

Newheiser, A.-K., Dunham, Y., Merrill, A., Hoosain, L., & Olson, K. R. (2014). Preference for high status predicts implicit outgroup bias among children from low-status groups. *Developmental Psychology*, *50*, 1081–1090. https://doi.org/10.1037/a0035054

Newheiser, A.-K., & Olson, K. R. (2012). White and Black American children's implicit intergroup bias. *Journal of Experimental Social Psychology*, *48*(1), 264–270. https://doi.org/10.1016/j.jesp.2011.08.011

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2021). Data Collection via Online Platforms: Challenges and Recommendations for Future Research. *Applied Psychology*, *70*(3), 1380–1402. https://doi.org/10.1111/apps.12302

Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology*, *18*(2), 297–317. https://doi.org/10.1348/026151000165706

Nijhof, Annabel. D., Brass, Marcel., & Wiersema, Jan. R. (2017). Spontaneous mentalizing in neurotypicals scoring high versus low on symptomatology of autism spectrum disorder. *Psychiatry Research*, *258*, 15–20. https://doi.org/10.1016/j.psychres.2017.09.060

Nishiyama, T., Suzuki, M., Adachi, K., Sumi, S., Okada, K., Kishino, H., Sakai, S., Kamio, Y., Kojima, M., Suzuki, S., & Kanne, S. M. (2014). Comprehensive Comparison of Self-administered Questionnaires for Measuring Quantitative Autistic Traits in Adults. *Journal of Autism and Developmental Disorders*, *44*(5), 993–1007. https://doi.org/10.1007/s10803-013-2020-7

Noguchi, Y., & Oizumi, R. (2018). Electric stimulation of the right temporo-parietal junction induces a task-specific effect in deceptive behaviors. *Neuroscience Research*, *128*, 33–39. https://doi.org/10.1016/j.neures.2017.07.004

Nordenbæk, C., Jørgensen, M., Kyvik, K. O., & Bilenberg, N. (2014). A Danish population-based twin study on autism spectrum disorders. *European Child & Adolescent Psychiatry*, *23*(1), 35–43. https://doi.org/10.1007/s00787-013-0419-5

Nuñez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *NeuroImage*, *25*(1), 267–277. https://doi.org/10.1016/j.neuroimage.2004.10.041

Oakley, B. F. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology*, *125*(6), 818–823. https://doi.org/10.1037/abn0000182

Ofen, N., Whitfield-Gabrieli, S., Chai, X. J., Schwarzlose, R. F., & Gabrieli, J. D. E. (2017). Neural correlates of deception: Lying about past events and personal beliefs. *Social Cognitive and Affective Neuroscience*, *12*(1), 116–127. https://doi.org/10.1093/scan/nsw151

O'Hair, D., Cody, M. J., Wang, X., & Chao, E. Y. (1990). Vocal stress and deception detection among Chinese. *Communication Quarterly*, *38*(2), 158–169. https://doi.org/10.1080/01463379009369751

Øksendal, E., Brandlistuen, R. E., Holte, A., & Wang, M. V. (2019). Peer-Victimization of Young Children With Developmental and Behavioral Difficulties—A Population-Based Study. *Journal of Pediatric Psychology*, *44*(5), 589–600. https://doi.org/10.1093/jpepsy/jsy112

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Oxford English Dictionary. (2022). In *Oxford Languages*. Oxford University Press. https://languages.oup.com/google-dictionary-en/

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive Function Deficits in High-Functioning Autistic Individuals: Relationship to Theory of Mind. *Journal of Child Psychology and Psychiatry*, *32*(7), 1081–1105. https://doi.org/10.1111/j.1469-7610.1991.tb00351.x

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Pantazi, M., Kissine, M., & Klein, O. (2018). The Power of the Truth Bias: False Information Affects Memory and Judgment Even in the Absence of Distraction. *Social Cognition*, *36*(2), 167–198. https://doi.org/10.1521/soco.2018.36.2.167

Parra-Frutos, I. (2013). Testing homogeneity of variances with unequal sample sizes. *Computational Statistics*, *28*(3), 1269–1297. https://doi.org/10.1007/s00180-012-0353-x

Partington, H., Chowdhury, I., Wu, R., Tojo, M., & White, S. J. (in prep.). *A Meeting of Minds: Is Mentalizing Absent in Autism or Dependent Upon Like-Mindedness?*

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, *134*(2), 270–300. https://doi.org/10.1037/0033-2909.134.2.270

Patterson, T. (2009). The Effect of Cognitive Load on Deception. *FIU Electronic Theses and Dissertations*. https://doi.org/10.25148/etd.FI09120811

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.

Pellicano, E., Dinsmore, A., & Charman, T. (2014). What should autism research focus upon? Community views and priorities from the United Kingdom. *Autism*, *18*(7), 756–770. https://doi.org/10.1177/1362361314529627

Perez-Zapata, D., Slaughter, V., & Henry, J. D. (2016). Cultural effects on mindreading. *Cognition*, *146*, 410–414. https://doi.org/10.1016/j.cognition.2015.10.018

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that…” attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Peskin, J. (1992). Ruse and representations: On children’s ability to conceal information. *Developmental Psychology*, *28*(1), 84–89. https://doi.org/10.1037/0012-1649.28.1.84

Peterson, C. C., Slaughter, V. P., & Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry*, *48*(12), 1243–1250. https://doi.org/10.1111/j.1469-7610.2007.01810.x

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Podda, J., Ansuini, C., Vastano, R., Cavallo, A., & Becchio, C. (2017). The heaviness of invisible objects: Predictive weight judgments from observed real and pantomimed grasps. *Cognition*, *168*, 140–145. https://doi.org/10.1016/j.cognition.2017.06.023

Pontillo, M., Tata, M. C., Averna, R., Demaria, F., Gargiullo, P., Guerrera, S., Pucciarini, M. L., Santonastaso, O., & Vicari, S. (2019). Peer Victimization and Onset of Social Anxiety Disorder in Children and Adolescents. *Brain Sciences*, *9*(6), 132. https://doi.org/10.3390/brainsci9060132

Pozzar, R., Hammer, M. J., Underhill-Blazey, M., Wright, A. A., Tulsky, J. A., Hong, F., Gundersen, D. A., & Berry, D. L. (2020). Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media:

Cross-Sectional Questionnaire. *Journal of Medical Internet Research*, *22*(10), e23021. https://doi.org/10.2196/23021

Prissé, B., & Jorrat, D. (2022). Lab vs online experiments: No differences. *Journal of Behavioral and Experimental Economics, 100*, 101910. https://doi.org/10.1016/j.socec.2022.101910

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review*, *27*(2), 224–260. https://doi.org/10.1016/j.dr.2007.02.001

Randall, P. (1996). *Adult Bullying: Perpetrators and Victims*. Routledge. https://doi.org/10.4324/9780203360255

Ranick, J., Persicke, A., Tarbox, J., & Kornack, J. A. (2013). Teaching children with autism to detect and respond to deceptive statements. *Research in Autism Spectrum Disorders*, *7*(4), 503–508. https://doi.org/10.1016/j.rasd.2012.12.001

Resume Builder. (2021, September 7). *1 in 3 Americans admit to lying on resume*. https://www.resumebuilder.com/1-in-3-americans-admit-to-lying-on-resume/

Rigby, K. (2005). Bullying in Schools and the Mental Health of Children. *Journal of Psychologists and Counsellors in Schools*, *15*(2), 195–208. https://doi.org/10.1375/ajgc.15.2.195

Riggio, R. E., Tucker, J., & Widaman, K. F. (1987). Verbal and nonverbal cues as mediators of deception ability. *Journal of Nonverbal Behavior*, *11*(3), 126–145. https://doi.org/10.1007/BF00990233

Risal, A., Manandhar, K., Linde, M., Koju, R., Steiner, T. J., & Holen, A. (2015). Reliability and Validity of a Nepali-language Version of the Hospital Anxiety and Depression Scale (HADS). *Kathmandu University Medical Journal*, *13*(2), Article 2. https://doi.org/10.3126/kumj.v13i2.16783

Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., Plomin, R., & Ronald, A. (2011). Evidence That Autistic Traits Show the Same Etiology in the General Population and at the Quantitative Extremes (5%, 2.5%, and 1%). *Archives of General Psychiatry*, *68*(11), 1113–1121. https://doi.org/10.1001/archgenpsychiatry.2011.119

Rødgaard, E.-M., Jensen, K., Miskowiak, K. W., & Mottron, L. (2022). Representativeness of autistic samples in studies recruiting through social media. *Autism Research*, *n/a*(n/a). https://doi.org/10.1002/aur.2777

Rodriguez, G., Drastal, K., & Hartley, S. L. (2021). Cross-lagged model of bullying victimization and mental health problems in children with autism in middle to older childhood. *Autism*, *25*(1), 90–101. https://doi.org/10.1177/1362361320947513

Roestorf, A., Gaigg, S. B., Williams, D. M., & Bowler, D. M. (2018). *Self-reports vs. Observer ratings of autistic traits: A comparison of their validity over the adult lifespan*. International Society for Autism Research (INSAR) Annual Meeting.

Rogge, N., & Janssen, J. (2019). The Economic Costs of Autism Spectrum Disorder: A Literature Review. *Journal of Autism and Developmental Disorders*, *49*(7), 2873–2900. https://doi.org/10.1007/s10803-019-04014-z

Rotenberg, K. J., Simourd, L., & Moore, D. (1989). Children's Use of a Verbal-Nonverbal Consistency Principle to Infer Truth and Lying. *Child Development*, *60*(2), 309–322. https://doi.org/10.2307/1130978

Rowley, E., Chandler, S., Baird, G., Simonoff, E., Pickles, A., Loucas, T., & Charman, T. (2012). The experience of friendship, victimization and bullying in children with an autism spectrum disorder: Associations with child characteristics and school placement. *Research in Autism Spectrum Disorders*, *6*(3), 1126–1134. https://doi.org/10.1016/j.rasd.2012.03.004

Rozmann, N., & Nahari, G. (2021). Credibility assessments of alibi accounts: The role of cultural intergroup bias. *Psychiatry, Psychology and Law*, *0*(0), 1–14. https://doi.org/10.1080/13218719.2021.1938274

Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin & Review*, *24*(4). https://doi.org/10.3758/s13423-016-1190-7

Russell, G., Stapley, S., Newlove-Delgado, T., Salmon, A., White, R., Warren, F., Pearson, A., & Ford, T. (2022). Time trends in autism diagnosis over 20 years: A UK population-based cohort study. *Journal of Child Psychology and Psychiatry*, *63*(6), 674–682. https://doi.org/10.1111/jcpp.13505

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. https://doi.org/10.1111/j.2044-835X.1991.tb00881.x

Rusticus, S., & Lovato, C. (2019). Impact of Sample Size and Variability on the Power and Type I Error Rates of Equivalence Tests: A Simulation Study. *Practical Assessment, Research, and Evaluation*, *19*(1). https://doi.org/10.7275/4s9m-4e81

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, *6*(1), 2. https://doi.org/10.1186/2040-2392-6-2

Rylaarsdam, L., & Guemez-Gamboa, A. (2019). Genetic Causes and Modifiers of Autism Spectrum Disorder. *Frontiers in Cellular Neuroscience*, *13*. https://www.frontiersin.org/article/10.3389/fncel.2019.00385

Saarinen, A., Jääskeläinen, I. P., Harjunen, V., Keltikangas-Järvinen, L., Jasinskaja-Lahti, I., & Ravaja, N. (2021). Neural basis of in-group bias and prejudices: A systematic meta-analysis. *Neuroscience & Biobehavioral Reviews*, *131*, 1214–1227. https://doi.org/10.1016/j.neubiorev.2021.10.027

Sai, L., Wu, H., Hu, X., & Fu, G. (2018). Telling a truth to deceive: Examining executive control and reward-related processes underlying interpersonal deception. *Brain and Cognition*, *125*, 149–156. https://doi.org/10.1016/j.bandc.2018.06.009

Sasson, N. J., & Bottema-Beutel, K. (2022). Studies of autistic traits in the general population are not studies of autism. *Autism*, *26*(4), 1007–1008. https://doi.org/10.1177/13623613211058515

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., & Grossman, R. B. (2017). Neurotypical Peers are Less Willing to Interact with Those with Autism based on Thin Slice Judgments. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep40700

Sasson, N. J., & Morrison, K. E. (2019). First impressions of adults with autism improve with diagnostic disclosure and increased autism knowledge of peers. *Autism*, *23*(1), 50–59. https://doi.org/10.1177/1362361317729526

Sato, W., & Uono, S. (2019). The atypical social brain network in autism: Advances in structural and functional MRI studies. *Current Opinion in Neurology*, *32*(4), 617–621. https://doi.org/10.1097/WCO.0000000000000713

Sattler, J. M. (1988). *Assessment of children, 3rd ed* (pp. xxviii, 995). Jerome M. Sattler.

Scheeren, A. M., de Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking theory of mind in high-functioning autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *54*(6), 628–635. https://doi.org/10.1111/jcpp.12007

Schmitt, L. M., White, S. P., Cook, E. H., Sweeney, J. A., & Mosconi, M. W. (2018). Cognitive mechanisms of inhibitory control deficits in autism spectrum disorder (ASD). *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(5), 586–595. https://doi.org/10.1111/jcpp.12837

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. https://doi.org/10.1016/j.cognition.2013.08.004

Schroeder, J. H., Cappadocia, M. C., Bebko, J. M., Pepler, D. J., & Weiss, J. A. (2014). Shedding Light on a Pervasive Problem: A Review of Research on Bullying Experiences Among Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *44*(7), 1520–1534. https://doi.org/10.1007/s10803-013-2011-8

Schubotz, R. I., & von Cramon, D. Y. (2009). The Case of Pretense: Observing Actions and Inferring Goals. *Journal of Cognitive Neuroscience*, *21*(4), 642–653. https://doi.org/10.1162/jocn.2009.21049

Schuwerk, T., Jarvers, I., Vuori, M., & Sodian, B. (2016). Implicit Mentalizing Persists beyond Early Childhood and Is Profoundly Impaired in Children with Autism Spectrum Condition. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.01696

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. https://doi.org/10.1098/rsos.172273

Semrad, M., Scott-Parker, B., & Nagel, M. (2019). Personality traits of a good liar: A systematic review of the literature. *Personality and Individual Differences*, *147*, 306–316. https://doi.org/10.1016/j.paid.2019.05.007

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind Eyes: An Absence of Spontaneous Theory of Mind in Asperger Syndrome. *Science*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Serious Science (Director). (2016, December 30). *Theory of Mind—Uta Frith*. https://www.youtube.com/watch?v=N6ylH-LYjOM

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The Prevalence of Lying in America: Three Studies of Self-Reported Lies. *Human Communication Research*, *36*(1), 2–25. https://doi.org/10.1111/j.1468-2958.2009.01366.x

Serota, K. B., Levine, T. R., & Docan-Morgan, T. (2021). Unpacking variation in lie prevalence: Prolific liars, bad lie days, or both? *Communication Monographs*, *0*(0), 1–25. https://doi.org/10.1080/03637751.2021.1985153

Shakoor, S., Jaffee, S. R., Bowes, L., Ouellet-Morin, I., Andreou, P., Happé, F., Moffitt, T. E., & Arseneault, L. (2012). A prospective longitudinal study of children's theory of mind and adolescent involvement in bullying: Theory of mind and adolescent involvement in bullying. *Journal of Child Psychology and Psychiatry*, *53*(3), 254–261. https://doi.org/10.1111/j.1469-7610.2011.02488.x

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., Sen, A., Singhal, N., Erickson, D., & Singh, N. C. (2010). Sounds of melody—Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45. https://doi.org/10.1016/j.neulet.2010.04.066

Shaw, H., & Lyons, M. (2017). Lie Detection Accuracy—The Role of Age and the Use of Emotions as a Reliable Cue. *Journal of Police and Criminal Psychology*, *32*(4), 300–304. https://doi.org/10.1007/s11896-016-9222-9

Sheppard, E., Pillai, D., Wong, G. T.-L., Ropar, D., & Mitchell, P. (2016). How Easy is it to Read the Minds of People with Autism Spectrum Disorder? *Journal of Autism and Developmental Disorders*, *46*(4), 1247–1254. https://doi.org/10.1007/s10803-015-2662-8

Shtayermman, O. (2007). Peer Victimization in Adolescents and Young Adults Diagnosed with Asperger's Syndrome: A Link to Depressive Symptomatology, Anxiety Symptomatology and Suicidal Ideation. *Issues in Comprehensive Pediatric Nursing*, *30*(3), 87–107. https://doi.org/10.1080/01460860701525089

Simmeborn Fleischer, A. (2012). Alienation and struggle: Everyday student-life of three male students with Asperger Syndrome. *Scandinavian Journal of Disability Research*, *14*(2), 177–194. https://doi.org/10.1080/15017419.2011.558236

Sip, K. E., Lynge, M., Wallentin, M., McGregor, W. B., Frith, C. D., & Roepstorff, A. (2010). The production and detection of deception in an interactive game. *Neuropsychologia*, *48*(12), 3619–3626. https://doi.org/10.1016/j.neuropsychologia.2010.08.013

Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: The scope and limits. *Trends in Cognitive Sciences*, *12*(2), 48–53. https://doi.org/10.1016/j.tics.2007.11.008

Smith, P. K. (2017). Bullying and Theory of Mind: A Review. *Current Psychiatry Reviews*, *13*(2), 90–95. https://doi.org/10.2174/1573400513666170502123214

Sodian, B. (1991). The development of deception in young children. *British Journal of Developmental Psychology*, *9*(1), 173–188. https://doi.org/10.1111/j.2044-835X.1991.tb00869.x

Sodian, B., & Frith, U. (1992). Deception and Sabotage in Autistic, Retarded and Normal Children. *Journal of Child Psychology and Psychiatry*, *33*(3), 591–605. https://doi.org/10.1111/j.1469-7610.1992.tb00893.x

Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early Deception and the Child's Theory of Mind: False Trails and Genuine Markers. *Child Development*, *62*(3), 468. https://doi.org/10.2307/1131124

Sofronoff, K., Dark, E., & Stone, V. (2011). Social vulnerability and bullying in children with Asperger syndrome. *Autism*, *15*(3), 355–372. https://doi.org/10.1177/1362361310365070

Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation Through Attribution of False Belief by 2-Year-Olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Sowden, S., Catmur, C., & Bird, G. (2018). F43. Lie Detection: How Autistic Traits Impact the Ability to Control Competing Representations of the Self and Others' Opinions. *Biological Psychiatry*, *83*(9), S254. https://doi.org/10.1016/j.biopsych.2018.02.656

Sowden, S., Wright, G. R. T., Banissy, M. J., Catmur, C., & Bird, G. (2015). Transcranial Current Stimulation of the Temporoparietal Junction Improves Lie Detection. *Current Biology*, *25*(18), 2447–2451. https://doi.org/10.1016/j.cub.2015.08.014

Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping*, *30*(11), 3704–3718. https://doi.org/10.1002/hbm.20800

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law*, *13*(1), 1–34. https://doi.org/10.1037/1076-8971.13.1.1

Sterzing, P. R., Shattuck, P. T., Narendorf, S. C., Wagner, M., & Cooper, B. P. (2012). Bullying Involvement and Autism Spectrum Disorders: Prevalence and Correlates of Bullying Involvement Among Adolescents With an Autism Spectrum Disorder. *Archives of Pediatrics & Adolescent Medicine*, *166*(11), 1058. https://doi.org/10.1001/archpediatrics.2012.790

Stewart, S. L. K., Wright, C., & Atherton, C. (2019). Deception Detection and Truth Detection Are Dependent on Different Cognitive and Emotional Traits: An Investigation of Emotional Intelligence, Theory of Mind, and Attention. *Personality and Social Psychology Bulletin*, *45*(5), 794–807. https://doi.org/10.1177/0146167218796795

Su, L., & Levine, M. (2016). Does "lie to me" lie to you? An evaluation of facial clues to high-stakes deception. *Computer Vision and Image Understanding*, *147*, 52–68. https://doi.org/10.1016/j.cviu.2016.01.009

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428–453. https://doi.org/10.1037/bul0000087

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months: Attribution of beliefs to a geometric shape. *British Journal of Developmental Psychology*, *30*(1), 30–44. https://doi.org/10.1111/j.2044-835X.2011.02046.x

Sutter, M. (2009). Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams. *The Economic Journal*, *119*(534), 47–60. https://doi.org/10.1111/j.1468-0297.2008.02205.x

Sutton, J., Smith, P. K., & Swettenham, J. (1999a). Bullying and 'Theory of Mind': A Critique of the 'Social Skills Deficit' View of Anti-Social Behaviour. *Social Development*, *8*(1), 117–127. https://doi.org/10.1111/1467-9507.00083

Sutton, J., Smith, P. K., & Swettenham, J. (1999b). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology*, *17*(3), 435–450. https://doi.org/10.1348/026151099165384

Sweeney, C. D., & Ceci, S. J. (2014). Deception detection, transmission, and modality in age and sex. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00590

Symes, W., & Humphrey, N. (2010). Peer-group indicators of social inclusion among pupils with autistic spectrum disorders (ASD) in mainstream secondary schools: A comparative study. *School Psychology International*, *31*(5), 478–494. https://doi.org/10.1177/0143034310382496

Talwar, V., & Crossman, A. (2011). Chapter 4 - From little white lies to filthy liars: The evolution of honesty and deception in young children. In J. B. Benson (Ed.), *Advances in Child Development and Behavior* (Vol. 40, pp. 139–179). JAI. https://doi.org/10.1016/B978-0-12-386491-8.00004-9

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology*, *43*(3), 804–810. https://doi.org/10.1037/0012-1649.43.3.804

Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of*

Behavioral Development, *26*(5), 436–444. https://doi.org/10.1080/01650250143000373

Talwar, V., & Lee, K. (2008). Social and Cognitive Correlates of Childrens Lying Behavior. *Child Development*, *79*(4), 866–881. https://doi.org/10.1111/j.1467-8624.2008.01164.x

Thurm, A., & Swedo, S. E. (2012). The importance of autism research. *Dialogues in Clinical Neuroscience*, *14*(3), 219–222.

Tilley, P., George, J. F., & Marett, K. (2005). Gender Differences in Deception and Its Detection Under Varying Electronic Media Conditions. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 24b–24b. https://doi.org/10.1109/HICSS.2005.284

Tonge, N. A., Rodebaugh, T. L., Fernandez, K. C., & Lim, M. H. (2016). Self-Reported Social Skills Impairment Explains Elevated Autistic Traits in Individuals with Generalized Social Anxiety Disorder. *Journal of Anxiety Disorders*, *38*, 31–36. https://doi.org/10.1016/j.janxdis.2015.12.005

Ung, D., McBride, N., Collier, A., Selles, R., Small, B., Phares, V., & Storch, E. (2016). The relationship between peer victimization and the psychological characteristics of youth with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *32*, 70–79. https://doi.org/10.1016/j.rasd.2016.09.002

Ure, A., Rose, V., Bernie, C., & Williams, K. (2018). Autism: One or many spectrums? *Journal of Paediatrics and Child Health*, *54*(10), 1068–1072. https://doi.org/10.1111/jpc.14176

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 & Adolescent Psychiatry*, *24*(11), 1399–1417. https://doi.org/10.1007/s00787-015-0689-1

van Roekel, E., Scholte, R. H. J., & Didden, R. (2010). Bullying Among Adolescents With Autism Spectrum Disorders: Prevalence and Perception. *Journal of Autism and Developmental Disorders*, *40*(1), 63–73. https://doi.org/10.1007/s10803-009-0832-2

Van Swol, L. M., Braun, M. T., & Kolb, M. R. (2015). Deception, Detection, Demeanor, and Truth Bias in Face-to-Face and Computer-Mediated Communication. *Communication Research*, *42*(8), 1116–1142. https://doi.org/10.1177/0093650213485785

Van Swol, L. M., Malhotra, D., & Braun, M. T. (2012). Deception and Its Detection: Effects of Monetary Incentives and Personal Relationship History. *Communication Research*, *39*(2), 217–238. https://doi.org/10.1177/0093650210396868

van Tiel, B., Deliens, G., Geelhand, P., Murillo Oosterwijk, A., & Kissine, M. (2021). Strategic Deception in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *51*(1), 255–266. https://doi.org/10.1007/s10803-020-04525-0

Vartanian, O., Kwantes, P., & Mandel, D. R. (2012). Lying in the scanner: Localized inhibition predicts lying skill. *Neuroscience Letters*, *529*(1), 18–22. https://doi.org/10.1016/j.neulet.2012.09.019

Veddum, L., Pedersen, H. L., Landert, A.-S. L., & Bliksted, V. (2019). Do patients with high-functioning autism have similar social cognitive deficits as patients with a chronic cause of schizophrenia? *Nordic Journal of Psychiatry*, *73*(1), 44–50. https://doi.org/10.1080/08039488.2018.1554697

Verigin, B. L., Meijer, E. H., Bogaard, G., & Vrij, A. (2019). Lie prevalence, lie characteristics and strategies of self-reported good liars. *PLOS ONE*, *14*(12), e0225566. https://doi.org/10.1371/journal.pone.0225566

Verkuil, B., Atasayi, S., & Molendijk, M. L. (2015). Workplace Bullying and Mental Health: A Meta-Analysis on Cross-Sectional and Longitudinal Data. *PLOS ONE*, *10*(8), e0135225. https://doi.org/10.1371/journal.pone.0135225

Verschuere, B., Köbis, N. C., Bereby-Meyer, Y., Rand, D., & Shalvi, S. (2018). Taxing the Brain to Uncover Lying? Meta-analyzing the Effect of Imposing Cognitive Load on the Reaction-Time Costs of Lying. *Journal of Applied Research in Memory and Cognition*, *7*(3), 462–469. https://doi.org/10.1016/j.jarmac.2018.04.005

Volkmar, F. R., & Mayes, L. C. (1990). Gaze behavior in autism. *Development and Psychopathology*, *2*(1), 61–69. https://doi.org/10.1017/S0954579400000596

Volz, K. G., Vogeley, K., Tittgemeyer, M., von Cramon, D. Y., & Sutter, M. (2015). The neural basis of deception in strategic interactions. *Frontiers in Behavioral Neuroscience*, *9*. https://www.frontiersin.org/articles/10.3389/fnbeh.2015.00027

Vrij, A. (1993). Credibility Judgments of Detectives: The Impact of Nonverbal Behavior, Social Skills, and Physical Characteristics on Impression Formation. *The Journal of Social Psychology*, *133*(5), 601–610. https://doi.org/10.1080/00224545.1993.9713915

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities / Aldert Vrij.* (2nd edition.). John Wiley.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, *10*(4), 141–142. https://doi.org/10.1016/j.tics.2006.02.003

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1), 1–21. https://doi.org/10.1111/lcrp.12088

Vrij, A., & Ganis, G. (2014). Chapter 7—Theories in Deception and Lie Detection. In D. C. Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility Assessment* (pp. 301–374). Academic Press. https://doi.org/10.1016/B978-0-12-394433-7.00007-5

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing Cognitive Load to Facilitate Lie Detection: The Benefit of Recalling an Event in Reverse Order. *Law and Human Behavior*, *32*(3), 253–265. https://doi.org/10.1007/s10979-007-9103-y

Vrij, A., Mann, S., Leal, S., & Fisher, R. (2010). 'Look into my eyes': Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law*, *16*(4), 327–348. https://doi.org/10.1080/10683160902740633

Wahyuningsih, H., & Novitasari, R. (2016). Theory of Mind and Victimization: A Preliminary Study on Indonesian Junior High School Students. *Procedia - Social and Behavioral Sciences*, *217*, 759–763. https://doi.org/10.1016/j.sbspro.2016.02.140

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, *17*(7), 755–774. https://doi.org/10.1002/acp.914

Wang, L. W., Berry-Kravis, E., & Hagerman, R. J. (2010). Fragile X: Leading the Way for Targeted Treatments in Autism. *Neurotherapeutics*, *7*(3), 264–274. https://doi.org/10.1016/j.nurt.2010.05.005

Wang, L., Zhu, L., & Wang, Z. (2017). Parental mind-mindedness but not false belief understanding predicts Hong Kong children's lie-telling behavior in a temptation resistance task. *Journal of Experimental Child Psychology*, *162*, 89–100. https://doi.org/10.1016/j.jecp.2017.04.023

Warland, H. (2014, July 14). *Theory of Mind in adults: Deception detection and reading facial expressions in the normal and autistic population—University of the Sunshine*

Coast, Queensland. University Research Conference: Communicate, Collaborate, Celebrate, Sunshine Coast, Australia. https://research.usc.edu.au/esploro/outputs/conferencePresentation/Theory-of-Mind-in-adults-Deception-detection-and-reading-facial-expressions-in-the-normal-and-autistic-population/99448770602621

Warrier, V., Greenberg, D. M., Weir, E., Buckingham, C., Smith, P., Lai, M.-C., Allison, C., & Baron-Cohen, S. (2020). Elevated rates of autism, other neurodevelopmental and psychiatric diagnoses, and autistic traits in transgender and gender-diverse individuals. *Nature Communications*, *11*(1), 3959. https://doi.org/10.1038/s41467-020-17794-1

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition*. https://doi.org/10.1037/t15169-000

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, *18*, 53–70. https://doi.org/10.1037/a0031607

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing Mentalizing Impairments in Autism. *Child Development*, *80*(4), 1097–1117. https://doi.org/10.1111/j.1467-8624.2009.01319.x

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of theory of mind for adults with autism. *Autism Research*, *4*(2), 149–154. https://doi.org/10.1002/aur.174

White, S. J., Frith, U., Rellecke, J., Al-Noor, Z., & Gilbert, S. J. (2014). Autistic adolescents show atypical activation of the brain′s mentalizing system even without a prior history of mentalizing problems. *Neuropsychologia*, *56*, 17–25. https://doi.org/10.1016/j.neuropsychologia.2013.12.013

Williams, D. M., Nicholson, T., Grainger, C., Lind, S. E., & Carruthers, P. (2018). Can you spot a liar? Deception, mindreading, and the case of autism spectrum disorder: Lie detection in ASD. *Autism Research*, *11*(8), 1129–1137. https://doi.org/10.1002/aur.1962

Williams, E. J., Bott, L. A., Patrick, J., & Lewis, M. B. (2013). Telling Lies: The Irrepressible Truth? *PLOS ONE*, *8*(4), e60713. https://doi.org/10.1371/journal.pone.0060713

Williams, G. L., Wharton, T., & Jagoe, C. (2021). Mutual (Mis)understanding: Reframing Autistic Pragmatic "Impairments" Using Relevance Theory. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.616664

Wilson, A. C. (2021). Do animated triangles reveal a marked difficulty among autistic people with reading minds? *Autism*, *25*(5), 1175–1186. https://doi.org/10.1177/1362361321989152

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening Adults for Asperger Syndrome Using the AQ: A Preliminary Study of its Diagnostic Validity in Clinical Practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335. https://doi.org/10.1007/s10803-005-3300-7

Wright, G. R. T., Berry, C. J., & Bird, G. (2012). “You can’t kid a kidder”: Association between production and detection of deception in an interactive deception task. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00087

Yi, L., Fan, Y., Li, J., Huang, D., Wang, X., Tan, W., Zou, X., & Lee, K. (2014). Distrust and retaliatory deception in children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *8*(12), 1741–1755. https://doi.org/10.1016/j.rasd.2014.09.006

Yin, L., Reuter, M., & Weber, B. (2016). Let the man choose what to do: Neural correlates of spontaneous lying and truth-telling. *Brain and Cognition*, *102*, 13–25. https://doi.org/10.1016/j.bandc.2015.11.007

Young, L., & Saxe, R. (2009). An fMRI Investigation of Spontaneous Mental State Inference for Moral Judgment. *Journal of Cognitive Neuroscience*, *21*(7), 1396–1405. https://doi.org/10.1162/jocn.2009.21137

Young, R. L., & Brewer, N. (2020). Brief Report: Perspective Taking Deficits, Autism Spectrum Disorder, and Allaying Police Officers’ Suspicions About Criminal Involvement. *Journal of Autism and Developmental Disorders*, *50*(6), 2234–2239. https://doi.org/10.1007/s10803-019-03968-4

Young, S. G. (2017). An outgroup advantage in discriminating between genuine and posed smiles. *Self and Identity*, *16*(3), 298–312. https://doi.org/10.1080/15298868.2016.1270851

Yuspeh, R. L., & Vanderploeg, R. D. (2000). Spot-the-Word: A Measure for Estimating Premorbid Intellectual Functioning. *Archives of Clinical Neuropsychology*, *15*(4), 319–326. https://doi.org/10.1093/arclin/15.4.319

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*(5), 778–790. https://doi.org/10.1002/aur.2696

Zeki, S., Goodenough, O. R., Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., & Ganesan, V. (2004). A cognitive neurobiological account of deception: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1451), 1755–1762. https://doi.org/10.1098/rstb.2004.1555

Zhang, M., Liu, T., Pelowski, M., & Yu, D. (2017). Gender difference in spontaneous deception: A hyperscanning study using functional near-infrared spectroscopy. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-06764-1

Zhao, C., Shang, S., Compton, A. M., Fu, G., & Sai, L. (2021). A Longitudinal Study of the Relations Between Theory of Mind, Executive Function, and Lying in Children. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.766891

Zheltyakova, M., Kireev, M., Korotkov, A., & Medvedev, S. (2020). Neural mechanisms of deception in a social context: An fMRI replication study. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-67721-z

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

Zloteanu, M., Krumhuber, E. G., & Richardson, D. C. (2021). Sitting in Judgment: How Body Posture Influences Deception Detection and Gazing Behavior. *Behavioral Sciences*, *11*(6), Article 6. https://doi.org/10.3390/bs11060085

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and Nonverbal Communication of Deception. In *Advances in Experimental Social Psychology* (Vol. 14, pp. 1–59). Elsevier. https://doi.org/10.1016/S0065-2601(08)60369-X

(Bottema-Beutel et al., 2021)

APPENDIX A: CHAPTER 4 ADDITIONAL ANALYSIS

One of the drawbacks of online collection of data was that there was no way concrete way to crosscheck the participants' diagnosis. While there were multiple safeguards to prevent wrongful grouping of participants, additional analyses were run on samples defined by the Autism Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, et al., 2001) scores.

Therefore, following the guidance laid by Woodbury-Smith et al. (2005), autistic participants who scored below 26 were excluded from the sample (*n* = 52), while non-autistic participants who scored above 32 (inclusive) were excluded (*n* = 23) for these additional analyses. The resulting sample will henceforth be called 'AQ-defined sample', as opposed to the 'original sample' that was analysed in the main body of the chapter.

A.1. Group matching

After exclusion, the final sample consisted of 567 non-autistic (NT) participants and 95 autistic (ASC) participants. As with the primary analysis, samples were mismatched in size, and were also not comparable on age and the number of second-language English speakers. Therefore, a similar group-matching method as the primary analyses was used. The flowchart in figure A.1 illustrates this further. The means and standard deviation of the pre-matching samples of each group and the matched group are presented in Table A.1.

Table A.1

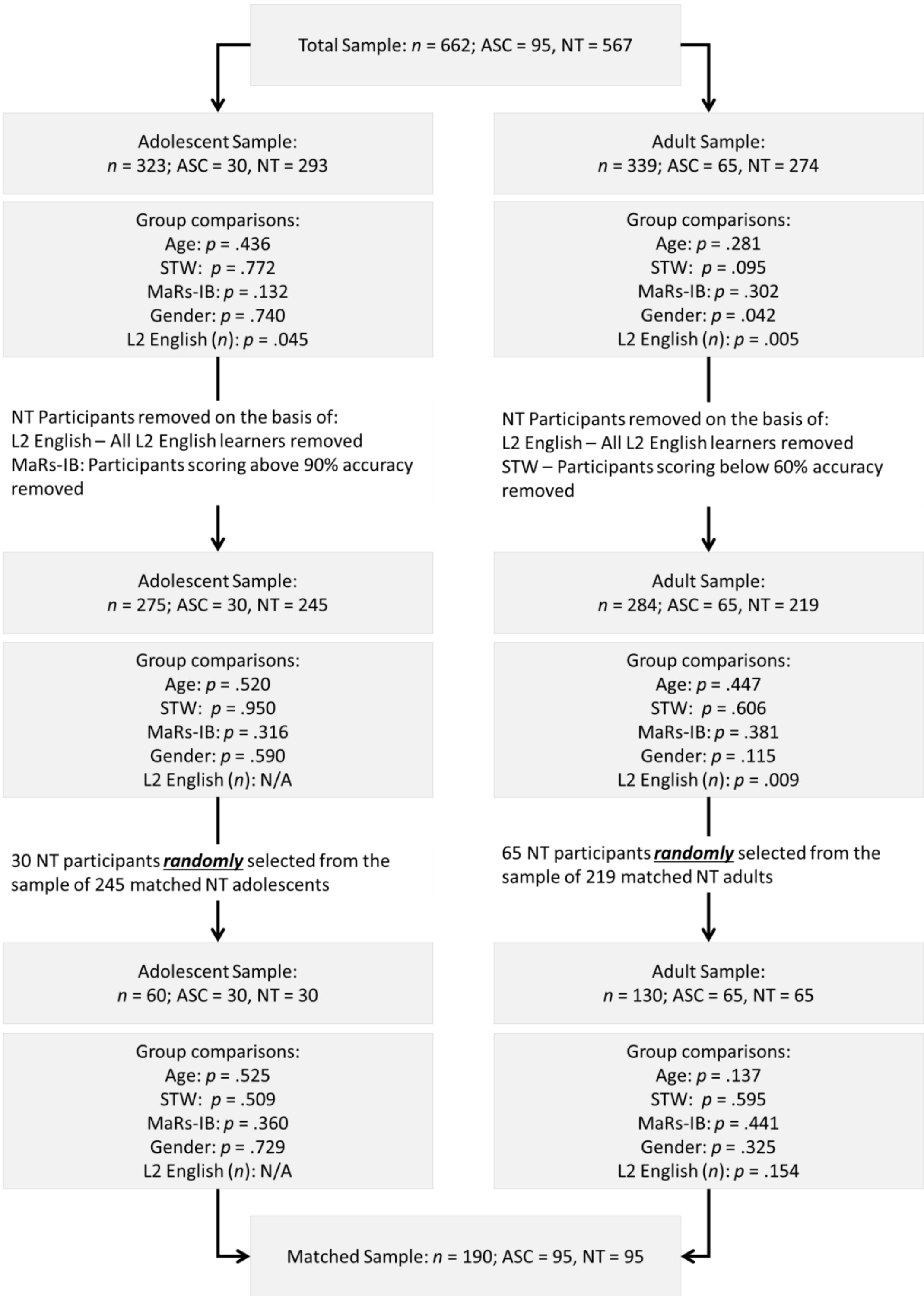
Sample demographics: Autistic sample compared to full non-autistic sample and matched non-autistic sample for additional analyses

	ASC (*n* = 95)		NT (*n* = 567)			Matched NT (*n* = 95)		
	Mean (*SD*)	Range	Mean (*SD*)	Range	*p*-value	Mean (*SD*)	Range	*p*-value
Age (years)	21.15 (5.84)	11-30	18.90 (5.62)	11-30	<.001	20.65 (5.31)	11-30	.542
STW	79.93 (10.78)	51.67-100	78.65 (9.97)	48.30-100	.254	79.74 (8.95)	60-98.33	.893
MaRs-IB	60.55 (18.81)	26.76-100	62.33 (18.41)	25.32-97.83	.386	60.32 (16.17)	27.50-100	.928
AQ	35.53 (6.55)	26-48	19.45 (6.11)	4-31	<.001	18.67 (6.16)	5-31	<.001
Gender	44 F; 44 M; 7 NB/O		297 F; 254 M; 16 NB/O		.065	47 F; 45 M; 3 NB/O		.425
L2 English learners	2		80		.001	0		.155

Note. STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), NB/O = Non-binary/Other/Prefer not to say; L2 English learners = Participants who are second language speakers of English

Figure A.1

Flowchart demonstrating the process of matching the ASC and NT groups on the basis of demographic information (age, gender, English fluency), background measures (verbal and non-verbal abilities), and sample size.



Note. STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), L2 English learners = Second language speakers of English

A.2. Mentalizing ability

Mentalizing ability was measured by the Frith-Happé animations (AT). A 3 x 2 mixed Analysis of Variance (ANOVA) was run with animation type (Random, GD, ToM) as the within subjects variable and diagnostic group (ASC / NT) as the between subjects factor (Figure A.2). No main effect of group was detected for the AQ-defined sample, F (1,188) = .129, p = .053, η_p^2 = .001, unlike when the same analysis was run on the original sample. However, a main effect for animation types was found, F (1.885,354.092) = 4.862, p = 0.010, η_p^2 = .025. Post-hoc t-tests showed that GD scores differed significantly from Random (p < .001), and that there was a marginal difference between participants' scores in the GD and ToM conditions (p = .057), but no difference between participants' performance in Random and ToM condition (p = .279). Participants scored highest in the AT-Random and lowest in the AT-GD condition. No interaction between animation type and group was observed (F [1.885, 354.092] = .179, p = .823, η_p^2 = .001), a deviation from the original results. See Table A.2 for all means and *SD*s.

Table A.2

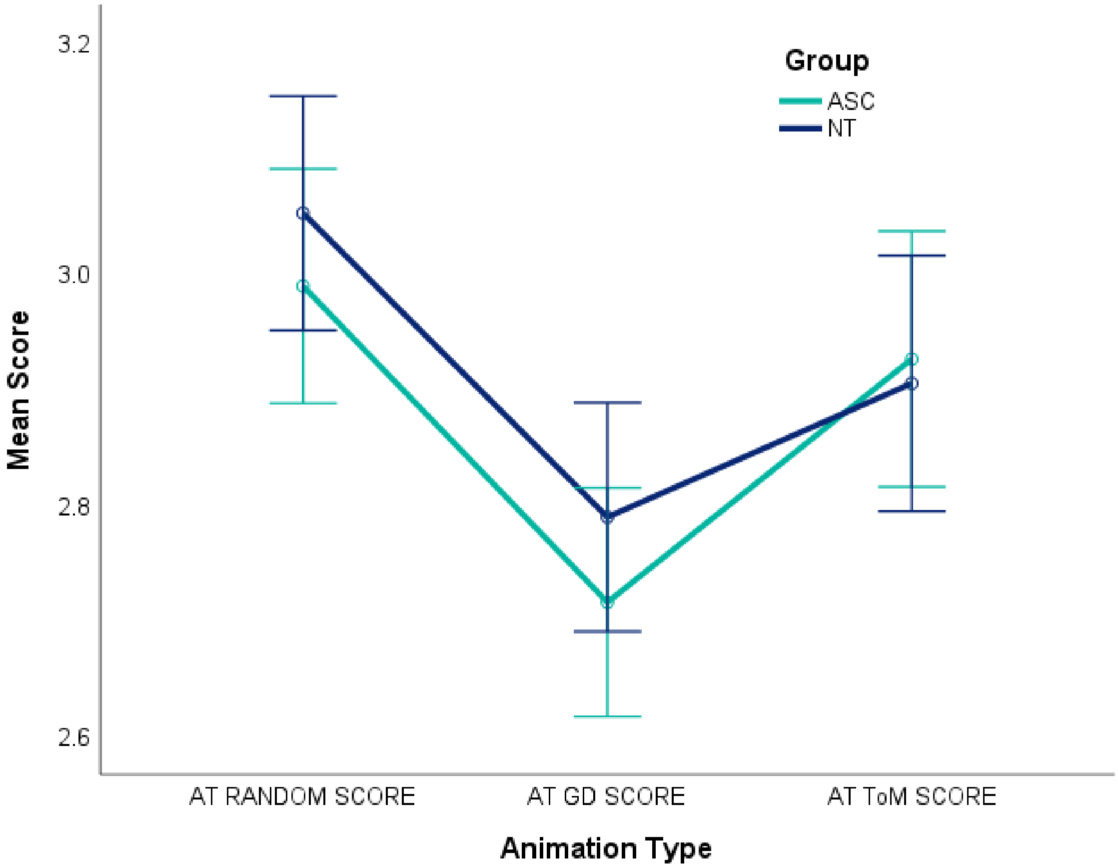
Descriptive statistics for the AT

	Across Groups		ASC (*n* = 95)		Matched-NT (*n* = 95)	
	M	*SD*	*M*	*SD*	*M*	*SD*
AT-Random	3.02	.99	2.99	.97	3.05	1.00
AT-GD	2.75	.93	2.72	1.00	2.79	.93
At-ToM	2.92	1.08	2.93	1.00	2.91	1.15
AT-Total	8.69	2.21	8.63	2.24	8.75	2.20

Note. n = sample size, *M* = mean, *SD* = standard deviation, AT = Frith-Happé animations, GD = Goal directed, ToM = Mentalizing

Figure A.2

Graphical illustration of participants' AT score



Note. Error bars represent one standard error.

A pair of Spearman's correlations was run to investigate the relationships of age with AT-Total scores and AT-ToM scores, separately for each receiver group. For both groups, age did not correlate with either AT-Total (ASC: r = .118, p = .253; NT: r = .086, p = .408) or AT-ToM (ASC: r = .050, p = .630; NT: r = .054, p = .604). However, the full NT group continued to correlate positively with age, similar to the original analyses; r = .102, p = .015.

The interaction between mean centred age and diagnosis, tested by a multiple linear regression, was not found to predict AT-Total score; β = .045, t (186) = .418, p = .677. A separate multiple linear regression for AT-ToM showed that the interaction was also non-significant for AT-ToM, β = .033, t (186) = .301, p = .763.

A.3. Deception detection ability

Deception detection ability was measured using three tasks – the Cheating Task (T. R. Levine, 2007–2011) and the novel tasks created for this project – the Weird Pictures (WP-3D) and the Two Truths and a Lie (TTL-3D).

A.3.1. Difference between groups

The differences between groups were re-analysed with the AQ-defined sample as with the original sample.

A.3.1.1. Cheating Task

Participants' Total CHR and that in the two transparency conditions were tested against chance (chance = 0). Overall CHR scores were above chance across groups and for each of the two groups separately in the AQ-defined sample. Similar results were observed for transparent CHR, while the opposite was true for non-transparent CHR, i.e. all participants (across groups and in each group separately) were below chance. All results were found to be significant at p = .001 level and showed medium to large effects sizes.

A 2 x 2 mixed Analysis of Variance (ANOVA) was run on the CHR to see if groups (ASC and NT-matched) differed on deception detection on the Cheating Task (Figure A.3), with group (autistic / non-autistic) as the between subject variable and transparency as the within subject variable. Table A.3 shows the means and standard deviations of both groups.

There was a significant main effect of transparency with participants scoring significantly higher in the transparent condition versus the non-transparent condition; F (1,188) = 341.48,

$p < .001$, $\eta_p^2 = .645$. However, as with the original results, there was no significant effect of group i.e. autistic receivers' deception detection overall was comparable to that of the matched non-autistic sample; F (1,88) = 0.643, $p = .424$, $\eta_p^2 = .003$. The interaction that was found in the original analysis was not found for the AQ-defined sample; F (1,188) = 1.069, $p = .303$, $\eta_p^2 = .006$. Even though the means score for the non-autistic group is higher than the autistic group, this difference is no longer significant, unlike the original findings.

Table A.3

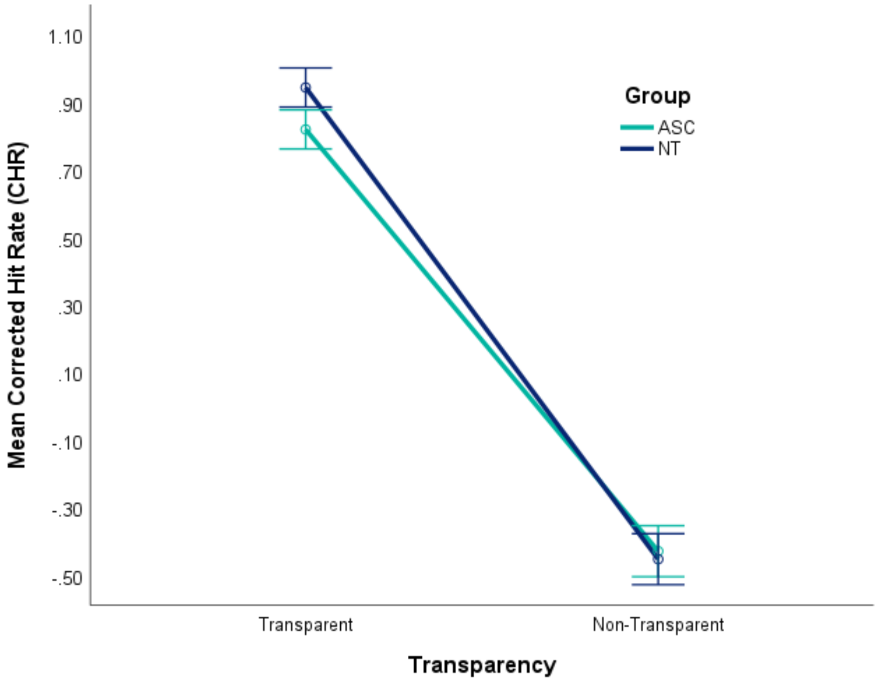
Descriptive statistics for autistic and (matched) non-autistic participants' CHR

Transparency	Group								
	Across groups			ASC			NT		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Total	190	0.29	.42	95	0.26	.45	95	0.32	0.39
Transparent	190	0.82	0.57	95	0.82	0.61	95	0.94	0.51
Non-transparent	190	-0.44	0.73	95	-0.43	0.75	95	-0.45	0.72

Note. ASC = Autistic sample, NT = matched autistic sample *n* = sample size, *M* = mean, *SD* = standard deviation

Figure A.3

Mean difference in performance on the Cheating Task between groups



Note. Error bars represent one standard error.

A.3.1.2. Novel Tasks

The AQ-defined non-autistic participants (*n* = 565) were above chance on both the WP-3D (*U* = 107487.00, *p* < .001) and the TTL-3D (*U* = 129926.00, *p* < .001), demonstrating that the tasks were working as expected.

Two independent samples Mann-Whitney U Tests were run to see if there was a difference in accuracy on the WP-3D and TTL-3D between the autistic participants and the (matched) non-autistic sample. Similar to the original findings, the ASC group performed slightly worse than the NT group on the WP-3D, however, this difference was no longer significant for the present sample; *U* = 4731.500, *z* = 0.741, *p* = .458, *r* = 0.05. The results from TTL-3D were identical to the original analysis, i.e. the two groups were comparable on their on their performance on the TTL-3D, *U* = 4542.000, *z* = 0.081, *p* = .936, *r* = 0.06. Mean, standard deviations and medians for both tasks of each group can be found in Table A.4.

Table A.4

Descriptive statistics for novel tasks for both receiver groups

	ASC			Matched NT		
	n	Median	*M* (*SD*)	n	Median	*M* (*SD*)
WP-3D	94	1.5	1.56 (1.06)	95	2	1.66 (0.92)
TTL-3D	95	2	1.86 (1.14)	95	2	1.86 (1.05)

Note. WP-3D = Weird Pictures (Development of Deception Detection), TTL-3D = Two Truths and a Lie (Development of Deception Detection)

A.3.2. Association analyses

Similar to the analyses conducted for the original sample, the deception detection performances of the two groups in the AQ-defined sample too were correlated with their age, verbal and non-verbal ability, autistic traits, and mentalizing ability. The groups were analysed separately due to the disparity in sample size, and in order to be able to compare and contrast the finding from this AQ-defined sample to those from the original sample. The Spearman's correlation coefficients for both groups are presented in table A.5.

For the ASC group, the overall trends observed for this sample were similar to the original sample, however certain associations that reached significance for the original sample did not do so here, while others that were not significant earlier were so now. Deception detection ability on the transparent condition of the Cheating Task showed positive associations only with verbal ability ($p < .001$) and AT-Total ($p = .017$), and not with age, non-verbal ability, or autistic traits like it did in the original sample, even though the correlation coefficients hint at similar trends. The WP-3D showed similar results in this sample as the original sample, as in that it only showed a positive association with autistic traits ($p = .003$). Performance on the

TTL-3D not only showed a positive association with AT-Total (*p* = .034) like the original sample, but also with verbal ability (*p* = .044) and autistic traits (*p* = .007).

For the NT group, the correlations for the AQ-defined sample showed the exact same trends as the original sample. Deception detection ability in the transparent condition of the Cheating Task positively correlated with age (p < .001) and verbal ability (*p* = .013), and also showed a week but significant negative correlation with atistic traits (*p* = .049). Performance on the WP-3D showed positive correlations with verbal ability (*p* = .032) and both the AT-Total (*p* = .001) and AT-ToM (*p* = .012). Lastly, scores on the TTL-3D showed a positive association only with verbal ability (*p* = .028).

(Rest of page left black intentionally)

Table A.5

Correlations between deception detection and predictor variables for ASC (above diagonal in grey) and NT (below diagonal) for AQ-defined sample

		1.	2.	3.	4.	5.	6.	7.	8.	9.	*M*	*SD*
1.	CHR-T	—	-0.021	0.135	0.144	.327***	0.135	0.193	.245*	0.071	.82	.61
2.	WP-3D	0.045	—	0.103	0.007	0.098	0.091	.302**	0.147	0.003	1.56	1.06
3.	TTL-3D	.166***	0.016	—	0.164	.207*	0.033	.277**	.218*	0.012	1.86	1.14
4.	Age	.213***	0.074	0.064	—	0.143	-0.011	0.180	0.118	0.050	21.15	5.84
5.	STW	.104*	.090*	.092*	0.045	—	.488***	.424***	.397***	0.170	79.93	10.78
6.	MaRs-IB	-0.026	0.078	0.029	-.148***	.277***	—	.274**	.435***	0.183	60.55	18.81
7.	AQ	-.083*	0.001	-0.050	-0.018	-.107*	-0.026	—	.344***	0.009	35.53	6.55
8.	AT-Total	0.056	.135**	0.064	.102*	.143***	.284***	-.095*	—	.627***	8.63	2.24
9.	AT-ToM	0.025	.106*	0.040	0.035	.089*	.253***	-.101*	.760***	—	2.93	1.00
	M	.90	1.59	1.92	18.90	78.65	62.33	19.45	8.88	2.89		
	SD	.57	1.02	1.05	5.62	9.97	18.41	6.11	2.15	1.06		

Note. Spearman's correlation coefficients for autistic receivers (*n* = 95, 94 for WP-3D) displayed above the diagonal; correlations for non-autistic receivers (*n* = 567, 565 for TTL-3D) displayed below the diagonal.

CHR-T = Corrected Hit Rate for Transparent Condition (Cheating Task), WP-3D = Weird Pictures (Development of Deception Detection), TTL-3D = Two Truths and a Lie (Development of Deception Detection), STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations Total Score, AT-ToM = Frith-Happé animations Mentalizing Score

p* < .05, ** *p* < .01, * *p* < .001 two tailed

A.3.3. Predicting deception detection ability

As with the original analyses, two five-block hierarchical linear regressions were run to predict deception detection – one for each group. All assumptions for multiple linear regression were met for both groups.

The same variables as the original analyses were entered in the exact same order to predict deception detection ability (CHR-T): demographic information – dummy coded gender variables, and English as primary language (L1) – was entered in the first block; proxy measures of verbal (STW) and non-verbal ability (MaRs-IB) were entered in the second block; age was entered in the third block; autistic traits was entered in the fourth block; mentalizing ability (AT-Total) was entered in the fifth block.

For ASC (Table A.6), Model 1, with just gender and English as primary language, did not predict any significant variance in deception detection ability, $F(3,91) = 1.149$, $p = .334$. Model 2 was significant ($F[5,89] = 2.693$, $p = .026$) and predicted an additional 9.5% of variance on the addition of verbal and non-verbal ability; verbal ability being a significant standalone predictor ($\beta = .31$, $p = .009$). Models 3 was significant; however, predicted only a megre 1% of additional variance in deception detection; $F[6,88] = 2.415$, $p = .033$. Models 4 ($F[7,87] = 2.049$, $p = .058$) and 5 ($F[8,86] = 2.008$, $p = .055$) were found to be only marginally signicant and while Model 4 did not predict any additional variance, Model 5 predicted an additional 1.6% of the variance. Overall, the final model accounted for 15.7% of the variance that was primarily driven by verbal ability, which emerged as the only significant unique predictor. These findings are partially similar to those derived from the models for the original sample where, for the ASC the model predicted around 11% variance which was mostly due to the predictor variables corresponding to verbal and non-verbal ability. Although, non-verbal ability does not seem to predict variance in this sample, verbal ability appears to be a much stronger predictor than earlier. Variables of age, autistic traits, and mentalizing, however, do not predict any variance in either sample.

For NT (Table A.7), Model 1, consisting of the dummy coded gender variables and English as primary language, did not predict any variance in deception detection ability F (3,563) = .914, $p = .434$. On addition of verbal and non-verbal ability, Model 2 was significant (F [5,561] = 3.153, $p = .008$) and explained 2.2% of the variance, with verbal ability being significant unique predictor ($\beta = .16$, $p < .001$). Model 3 was also significant (F [6,560] = 6.629, $p < .001$), predicting an additional 3.9% of variance with the addition of age ($\beta = .20$, $p < .001$). Model 4 and 5, contributing a megre .3% and .4% additional variance respectively, were also significant (Model 4: F [7,569] = 5.950, $p < .001$; Model 5: F [6,58] = 5.538, $p < .001$), however neither autistic traits nor mentalizing were unique predictors. Verbal ability and age continued to be strong significant unique predictors in the final model, which predicted a total of 7.4% variance in the deception detection. These results are starkly similar to those of the original sample, where the final model predicted 7.6% of the variance in deception detection, with verbal ability and age as the most significant unique predictors.

(Rest of page intentionally left blank)

Table A.6

Summary of Hierarchical Regression Analysis predicting deception detection for autistic receivers in the AQ-defined sample

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Female	-.30	.26	-.24	-.20	.25	-.16	-.18	.25	-.15	-.18	.26	-.15	-.17	.26	-.14
Male	-.12	.25	-.10	.03	.25	.03	.04	.25	.03	.03	.26	.02	.06	.26	.05
Eng-L1	.56	.45	.13	.36	.44	.09	.40	.44	.09	.40	.44	.09	.33	.44	.08
STW				.02	.01	.31**	.02	.01	.30*	.02	.01	.30*	.02	.01	.28*
MaRs-IB				.00	.00	.02	.00	.00	.03	.00	.00	.03	.00	.00	-.01
Age							.01	.01	.10	.01	.01	.10	.01	.01	.08
AQ										.00	.01	-.01	.00	.01	-.03
AT-Total													.04	.03	.15
R^2	.036			.131			.141			.142			.157		
ΔR^2	.036			.095			.010			.000			.016		
ΔF	1.15			4.86			1.02			.01			1.62		

Note. $n = 95$. Eng-L1 = English as primary language, STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations total score (mentalizing).
*$p < .05$, ** $p < .01$ two tailed.

Table A.7

Summary of Hierarchical Regression Analysis predicting deception detection for non-autistic receivers in the AQ-defined sample

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Female	.16	.15	.14	.19	.15	.17	.19	.14	.17	.17	.14	.15	.17	.14	.15
Male	.19	.15	.16	.19	.15	.17	.17	.14	.15	.15	.14	.13	.15	.14	.13
Eng-L1	-.07	.07	-.04	-.08	.07	-.05	-.06	.07	-.04	-.06	.07	-.04	-.06	.07	-.04
STW				.01	.00	.16***	.01	.00	.15***	.01	.00	.14**	.01	.00	.14**
MaRs-IB				.00	.00	-.07	.00	.00	-.03	.00	.00	-.03	.00	.00	-.05
Age							.02	.00	.20***	.02	.00	.20***	.02	.00	.19***
AQ										-.01	.00	-.06	.00	.00	-.05
AT-Total													.02	.01	.07
R^2	.005			.027			.066			.069			.074		
ΔR^2	.005			.022			.039			.003			.004		
ΔF	.91			6.49			23.38			1.82			2.54		

Note. n = 567. Eng-L1 = English as primary language, STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations total score (mentalizing).

** $p < .01$, *** $p < .001$ two tailed.

A.3.4. Deception detection and peer victimization

The autistic group (*M* = 34.45, *SD* = 11.88) in this AQ-defined sample had a higher mean score than the matched non-autistic sample (*M* = 32.86, *SD* = 10.86), however, unlike the original sample this difference was not significant, t (188) = .978, p = .329, d = .142.

To explore if mentalizing and deception detection predicted peer-victimization, robust multiple regressions for each diagnostic group in the AQ-defined sample were conducted as the assumption of normality of residuals was violated. For each group, participants' mentalizing scores. i.e., the AT-Total, and deception detection scores, i.e. CHR in transparent condition (CHR-T) were entered into the model as regressors along with background factors such as gender, age, verbal and non-verbal ability. All regression coefficients can be found in table A.8.

For the ASC group (n = 95), the overall model was significant (F [8,86] = 5.609, $p < .001$) and predicted about 34.3% percent of variance in peer-victimization scores, compared to the 36.3% in the original sample. Surprisingly, where the MarRs-IB stood out as the strongest unique predictor of peer-victimization in the ASC group in the original sample, that was no longer the case for the AQ-adjusted sample here. Instead verbal ability emerged as a marginally significant predictor (β = - .24, p = .079), suggesting that it is likely that those with a higher verbal ability may be victimized less. Age was also a unique predictor (β = .18, p = .036) indicating that for the ASC participants in this AQ-defined sample, experiences of peer-victimization increased as they grew older, just like it was seen for the original sample. Similar to earlier results from the original sample, mentalizing ability, i.e. AT-Total, here too was not found to be a significant predictor. Deception detection, i.e., CHR-T, was an significant unique predictor (β = -.19, p = .030), showing that the better the deception detection ability, the lower the peer-victimization reported, albeit this significance was weaker than that in the original sample.

In the NT group (n = 567), the overall model was also significant (F [8,558] = 6.986, p < .001) and predicted 9.1% of variance in reported peer-victimization, compared to the 9.3% in the original sample. Similar to the original sample, and unlike the ASC group here, non-verbal ability emerged as significant unique predictor (β = -.14, p = .007). Age was a unique predictor in the NT group as well (β = -.09, p = .028), but as with the original sample, the direction of this relationship was opposite to what was observed in the ASC group: reported experience of peer-victimization decreased with age for the NT participants. Unlike the ASC group, AT-Total emerged as the strongest unique predictor of variance in the NT group (β = -.16, p < .001). In this too, the findings were comparable to those of the original sample. Deception detection, i.e. CHR-T, similar to the original results, remained a marginally significant unique predictor (β = -.08, p = .077), demonstrating that higher deception detection accuracy corresponds to lower reported peer-victimization. The only difference between the findings of this sample and the original sample is that autistic traits emerged as a significant predictor here (β = .09, p = .022), which suggest that higher the autistic traits, higher the victimization experienced by the non-autistic individual.

(Rest of page intentionally left blank)

Table A.8

Summary of robust multiple regression analysis predicting peer-victimization

	ASC				NT			
	B	*SE B*	β	*p*	*B*	*SE B*	β	*p*
Female	3.92	4.34	.17	.353	-2.77	2.83	-.13	.319
Male	5.83	4.26	.25	.152	-1.69	2.82	-.08	.541
Age	.37	.17	.18	**.036**	-.16	.07	-.09	**.028**
STW	-.26	.14	-.24	**.079**	.00	.05	.00	.981
MaRs-IB	-.12	.08	-.18	.123	-.08	.03	-.14	**.007**
AQ	-.20	.19	-.11	.278	.14	.06	.09	**.022**
AT-Total	-.34	.59	-.06	.574	-.76	.23	-.16	**<.001**
CHR-T	-3.72	1.69	-.19	**.030**	-1.41	.79	-.08	**.077**

Note. Unstandardized coefficient and error bias-corrected and accelerated bootstrapped at 2000 samples. Standardised coefficient not bootstrapped.

Female and Male are dummy coded gender variables, non-binary / other is coded as 0 for both variables. STW = Spot the Word (verbal ability), MaRs-IB = Matrix Reasoning Item Bank (non-verbal ability), AQ = Autistic Quotient (autistic traits), AT-Total = Frith-Happé animations total score (mentalizing), CHR-T = Corrected hit rate for transparent condition on Cheating task (deception). Bold indicates significant or marginally significate results.

A.3.5. Deception detection, peer-victimization, and mental health

Similar to the original sample, an independent samples t-test confirmed that autistic individuals in the AQ-defined sample (*M* = 18.89, *SD* = 7.95) reported higher psychological distress than matched non-autistic individuals (*M* = 13.60, *SD* = 6.51), t (180.96) = 5.022, $p < .001$, $d = .729$.

The hypothesis regarding the relationship between deception detection, peer-victimization, and mental health (H_{10}) did not include specific predictions for the two diagnostic groups and the corresponding mediation analysis was conducted on the full original sample, without autistic traits or autism diagnosis being a factor / variable of interest. In other words, the sample was not segregated into autistic and non-autistic groups for this research question and therefore, the concerns regarding the overlap in autistic traits in the two groups and potential wrongful grouping of participants is not a concern for this specific hypothesis / analysis. Hence, the mediation analysis was not re-run for the AQ-defined sample.

APPENDIX B: CHAPTER 5 ADDITIONAL ANALYSIS I

As the non-autistic receivers in the WP-DEP did not perform above chance, any further findings from the DEP were deemed uninterpretable. Even so, the analyses for this task is documented here.

B.1. Main effect of receiver diagnosis

A Mann-Whitney U-test was run to check if the receiver groups differed on how well they judged lies, regardless of the sender labels. In the WP-DEP, scored out of 4, autistic receivers (M = 1.65, SD = 1.05, Mdn = 2) were not significantly worse at detecting deception than non-autistic receivers (M = 1.75, SD = 0.93, Mdn = 2) were, U = 4912.00, z = -0.841, p = .401, r = -.05.

B.2. Main effect of congruency

A Wilcoxon signed-rank test was run to see if lie detection differed depending on congruency between sender label and receiver diagnosis. Receiver's lie detection in WP-DEP did not significantly differ between the congruent (M = 0.85, SD = 0.69, Mdn = 1) and incongruent conditions (M = 0.85, SD = 0.76, Mdn = 1), T = 3554.50, z = -0.044, p = .965, r = -.003.

B.3. Interaction between receiver diagnosis and congruency

To examine if congruency between sender labels and receiver diagnosis affected performance differently for autistic and non-autistic receivers, an 'interaction score' was calculated by subtracting the receivers' scores in incongruent conditions from their scores in congruent condition, which could range from -2 to +2. A higher score denoted better lie detection for congruent senders. A significant difference between groups in their interaction scores would indicate that congruency affected lie detection performance differently in each group. Descriptive statistics for the difference scores / interaction scores on both tasks for each group can be found in Table 5.7.

A Mann-Whitney U-test demonstrated that there was no difference between the autistic receivers (*M* = 0.04, *SD* = 0.96, *Mdn* = 0) and the non-autistic receivers (*M* = -0.03, *SD* = 0.92, *Mdn* = 0) in how congruency affected their performance in the WP-DEP, *U* = 5498.00, *z* = 0.609, *p* = .543, *r* = .04

These results must be interpreted in the context of the main effect of congruency. As there was no effect of congruency – i.e. across groups, receivers did not show any difference in their deception detection abilities in either congruent or congruent conditions – the interaction demonstrates that neither group performed better in any condition. In other words, the results from the WP-DEP show that both groups performed comparably in all conditions.

Appendix C: Chapter 5 Additional Analysis II

As with Chapter 4, a concern with the online nature of data collection in the Double Empathy Problem studies in Chapter 5 was that participants may have been wrongly grouped, as the grouping was based solely on participant report without any means to crosscheck. Although there were multiple safeguards in this study, along with Prolific's own strict criteria check during the induction process, additional analyses were run on samples defined by the Autism Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, et al., 2001) scores. As with the additional analysis for Chapter 4 (Appendix B), the Woodbury-Smith et al. (2005) cut-offs for clinically diagnosed / referred sample (threshold value 26) and general population (threshold value 32) were used. Therefore, following these guidance, autistic participants who scored below 26 were excluded from the sample (*n* = 25), while non-autistic participants who scored above 32 (inclusive) were excluded (*n* = 5) for these additional analyses. The resulting sample will henceforth be called 'AQ-defined sample', as opposed to the 'original sample' that was analysed in the main body of the chapter. The resulting sample will henceforth be called 'AQ-defined sample', as opposed to the 'original sample' that was analysed in the main body of the chapter.

C.1. Participant information and group matching

After the exclusion, the final sample consisted of 78 autistic (ASC) participants and 97 non-autistic (NT) participants. Table C.1 provides a group-wise breakdown of participant demographic details and scores on background measures, along with information on how the two groups in this AQ-defined sample matched on these criteria.

Table C.1

Sample demographics and group matching for the AQ-defined sample

	ASC (*n* = 78)		NT (*n* = 97)		*t*-test
	Mean (*SD*)	Range	Mean (*SD*)	Range	*p*-value
Age (years)	36 (*10.16*)	18-57	34.04 (*9.93*)	19-60	.201
Verbal ability[1]	84.36 (*10.13*)	51.67-98.33	83.45 (*9.54*)	60-100	.545
Non-verbal ability[2]	61.50 (*16.00*)	30-92.59	58.71 (*16.08*)	30.65-89.74	.268
Autistic traits[3]	38.05 (*6.08*)	26-50	17.97 (*6.95*)	5-32	<.001
Gender	35 F; 37 M; 6 NB		51 F, 46 M		.018
Handedness	8 L; 62 R; 8 AMB		10 L; 85 R; 2 AMB		.066
L2 English learners	3		14		.019

Note. NB = Non-binary; AMB = Ambidextrous; L2 English learners = Participants who are second language speakers of English.

[1]Verbal ability as measured by accuracy percentage on Spot The Word (STW, Baddeley et al., 1993; adapted for online testing as seen in Livingston et al., 2021). Note that three autistic participants from Wave 1 had an accuracy below 60% but were still included in the sample. This was done because the Wave 1 exclusion criteria was based on above chance performance (50%) as a data quality (attention) check while Wave 2 was set to auto-exclude participants below 60% as the STW in Wave 2 also acted like a pre-screen for verbal ability. As the groups are matched on verbal ability, this was not considered an issue.

[2] Non-verbal ability as measured by accuracy percentage on the Matrix Reasoning Item Bank (MaRs-IB, Chierchia et al., 2019)

[3]Autistic traits as measured by the Autism Spectrum Quotient (AQ, Baron-Cohen, Wheelwright, Skinner, et al., 2001)

As evident from Table X.x , a series of t-tests confirmed that two groups were comparable on verbal and non-verbal ability, as well as age, and only differed in autistic traits: autistic participants reported significantly higher autistic traits as measured by the AQ (Baron-Cohen, Wheelwright, Skinner, et al., 2001). Unfortunately, the two groups did not have the same gender breakdown – there were six non-binary autistic participants, while none of non-autistic participants identified as non-binary. However, since gender was not observed to have an effect on deception detection in Chapter 4 and there is an established link between gender diversity and autism (Warrier et al., 2020), therefore making the sample more representative, no further steps were taken to match the two groups on gender.

C.2. Group Identification

A mixed 2 (receiver group: ASC / NT) x 2 (group membership: in-group / out-group) Analysis of Variance (ANOVA) demonstrated that receivers identified significantly more with their in-group compared to their out-group; F (1,173) = 423.941, $p < 0.001$, $\eta_p^2 = .710$. However, receiver group also significantly interacted with group membership (F [1, 173] = 11.024, $p = .001$, $\eta_p^2 = .060$): Autistic receivers identified with their in-group significantly less than non-autistic receivers (t [173] = -2.059, $p = .041$, $d = -.313$), while also identifying with their out-group much more than non-autistic receivers did (t [173] = 3.564, $p < .001$, $d = .542$). Means and standard deviations of how much receivers identified with their in- and out-groups are in Table C.2.

Table C.2

Means and standard deviations of Group-IDQ for the AQ-defined sample

	All receivers		Autistic receivers		Non-autistic receivers	
	M	*SD*	*M*	*SD*	*M*	*SD*
In-group identification	5.31	1.28	5.10	1.28	5.49	1.27
Out-group identification	2.35	1.03	2.65	0.99	2.11	1.28

Note. *M* = Mean, *SD* = Standard deviation. A mean score of 1 indicates low identification and a mean score of 7 indicates high identification. A mean score of 3.5 would be a non-biased response.

C.3. Investigating the double empathy problem: Box lifts

Similar to the analyses conducted for the original sample, a 2 x 2 x 2 mixed ANOVA was conducted on the deception rating with receiver **group** (ASC / NT) as the between-subject variable, and **authenticity** of box lift (Real / Pretend) and **congruency** between receiver and sender neurotype (Congruent / Incongruent) as the within-subject variables. The means and standard deviations of all conditions and groups are presented in Table C.3.

C.3.1. Main effects

Akin to the findings from the original sample, there was no significant effect of group (F [1, 173] = 0.069, p = .793, η_p^2 = .000) or congruency (F [1, 173] = 1.629, p = .204, η_p^2 = .009). A significant main effect of authenticity was found (F [1, 173] = 580.002, p < .001, η_p^2 = .746): participants in the AQ-defined sample rated Pretend videos as being more fake than Real videos, just like the participants in the original sample.

Table C.3

Means and standard deviations of video ratings in all conditions and groups

Authenticity	Congruency	Group					
		ASC		NT		Across groups	
		M	*SD*	*M*	*SD*	*M*	*SD*
Across Authenticity	Across Congruency	2.78	0.57	2.80	0.50	2.79	0.53
	Congruent	2.88	0.60	2.65	0.53	2.76	0.57
	Incongruent	2.67	0.60	2.94	0.55	2.82	0.59
Real	Across Congruency	2.38	0.61	2.38	0.58	2.38	0.59
	Congruent	2.39	0.63	2.29	0.61	2.34	0.62
	Incongruent	2.36	0.65	2.47	0.64	2.42	0.65
Pretend	Across Congruency	3.18	0.64	3.22	0.53	3.20	0.58
	Congruent	3.38	0.69	3.02	0.58	3.18	0.65
	Incongruent	2.98	0.69	3.41	0.61	3.22	0.68

Note. ASC = autistic receivers, NT = non-autistic receivers, *M* = mean, *SD* = standard deviation

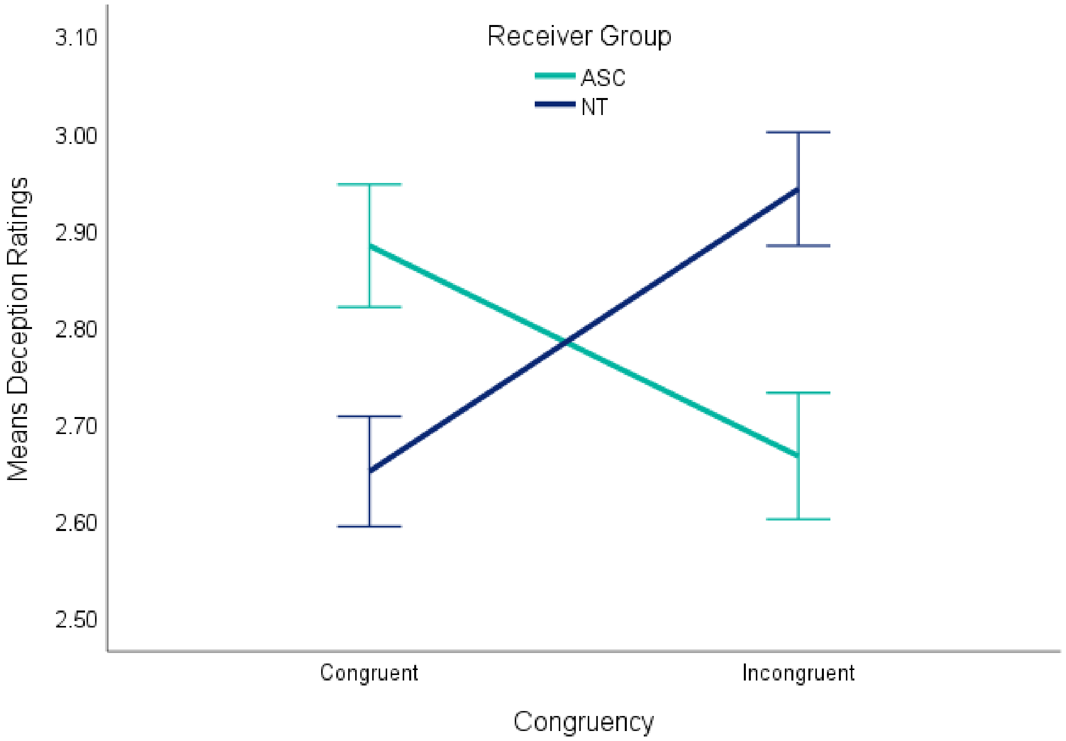
C.3.2. Interactions

There was no interaction between Congruency and Authenticity; F (1, 173) = 2.315, p = .130, η_p^2 = .013. Receivers in the AQ-defined sample discriminated between real and pretend lifts similarly for in-group and out-group senders, same as the receivers in the original sample.

There was a significant interaction between Congruence and Group; F (1, 173) = 76.163, p < .001, η_p^2 = .306 (Figure C.1). Post hoc paired *t*-tests showed that autistic participants rated videos as being significantly more fake in the congruent condition (autistic senders) than in the incongruent condition (non-autistic senders), t (77) = 5.14, p < .001, d = .582. Conversely, non-autistic participants rated videos as being significantly more fake in the incongruent condition (autistic senders) than in the congruent condition (non-autistic senders), t (96) = -7.34, p < .001, d = -.745. These results are identical to those of the original sample.

Figure C.1

Interaction between Congruency and Receiver Group



Note. The scale of the graph has been adjusted to illustrate the differences between categories better; the Y-axis does not start at zero. Error bars represent one standard error.

The interaction between Authenticity and Group for the AQ-defined sample was not significant (*F* [1, 173] = .221, $p = .639$, $\eta_p^2 = .001$), suggesting that there was no difference between the groups on their ability to distinguish between the real and pretend lifts. This is a departure from the findings from the original sample, where autistic receivers distinguished between rela and pretend lifts less than non-autistic receivers did.

There was a significant three-way interaction between Group, Authenticity, and Congruency, F (1, 173) = 37.049, $p < .001$, $\eta_p^2 = .176$.

C.3.3. Post-hoc ANOVAs

As with the original sample, this three-way interaction in the AQ-defined sample was investigated further by conducting two separate 2 x 2 repeated measures ANOVAs were conducted, one each for autistic receivers (ASC) and non-autistic receivers (NT), with Congruency (Congruent / Incongruent) and Authenticity (Real / Pretend) as the within subjects factor.

C.3.3.1. Congruency and deception detection in autistic receivers

For the autistic receivers, there was significant main effect of Authenticity; F (1, 77) = 208.073, $p < .001$, $\eta_p^2 = .730$. Autistic receivers rated pretend lifts as being significantly more fake than real lifts, indicating intact ability to detect deception. There was also significant main effect of congruency, F (1, 77) = 26.412, $p < .001$, $\eta_p^2 = .255$. This indicated that autistic receivers rated their in-group senders (i.e. autistic senders) much higher on deception, i.e. rated them as being more fake across trials, than their out-group senders (i.e. non-autistic senders), regardless of their actual veracity.

There was a significant interaction between Authenticity and Congruency for autistic receivers; F (1, 77) = 31.480, p < .001, η_p^2 = .290 (Figure C.2a). Paired t-tests demonstrated that both in-group and out-group senders were rated comparably in the real lifts (t [77] = .795, p = .215, d = -.133), but autistic receivers rated in-group senders more fake than out-group senders for the pretend lifts (t [77] = 6.676, p < .001, d = 756). This showed that autistic receivers were better at detecting deception in the in-group than the out-group but were more likely to identify other autistic individuals as being deceptive. Overall, these results were identical to that found for the original sample

C.3.3.2. Congruency and deception detection in non-autistic receivers

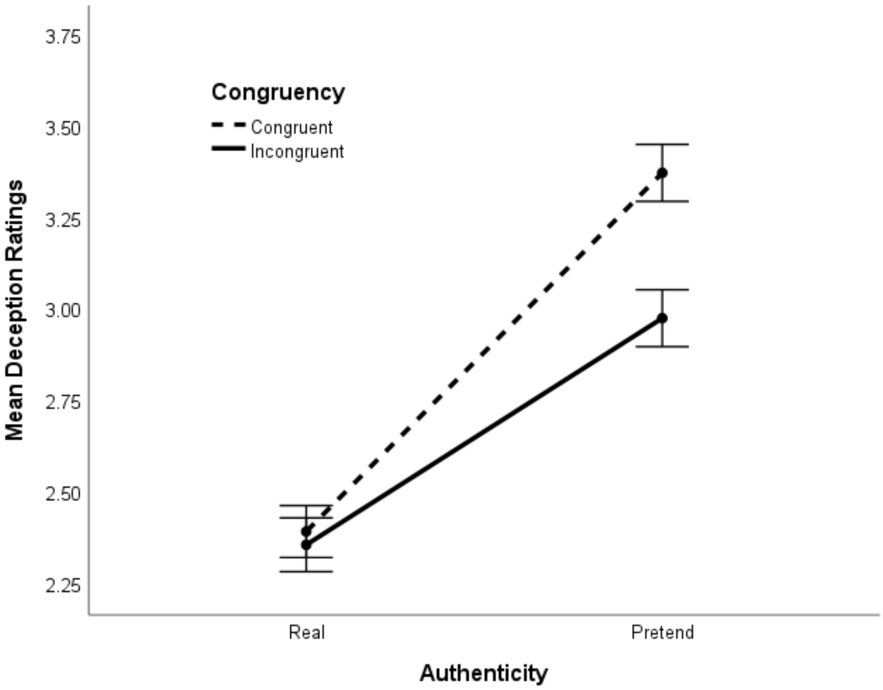
There was a significant main effect of Authenticity for the non-autistic receivers; F (1, 96) = 310.839, p < .001, η_p^2 = .764. Participants rated pretend videos significant higher on deception than real videos. The main effect of congruency was also significant, F (1, 96) = 53.882, p < .001, η_p^2 = .359: the non-autistic group rated out-group (i.e. autistic senders) videos much higher on deception in comparison to the in-group (i.e. non-autistic senders).

The interaction between Authenticity and Congruency was also found to be significant for the non-autistic receivers, F (1, 96) = 10.282, p = .002, η_p^2 = .097 (Figure C.2b). Post-hoc paired t-tests revealed that the non-autistic group rated out-group senders as being more fake than in-group senders for both real lifts (t [96] = -3.836 p < .001, d = -.388) and pretend lifts (t [96] = -7.123, p < .001, d = -.723). The difference between incongruent and congruent, however, was significantly more pronounced for the pretend lifts (M = -.40, SD = 0.55) than real lifts (M = -.18, SD = 0.47); t (96) = 3.207, p < .001, d = .326. This indicates that, non-autistic receivers were better at deception detection for the out-group than they were for the in-group. As with the AQ-defined ASC sample above, these results for the AQ-defined NT sample were identical to the findings of the original NT sample.

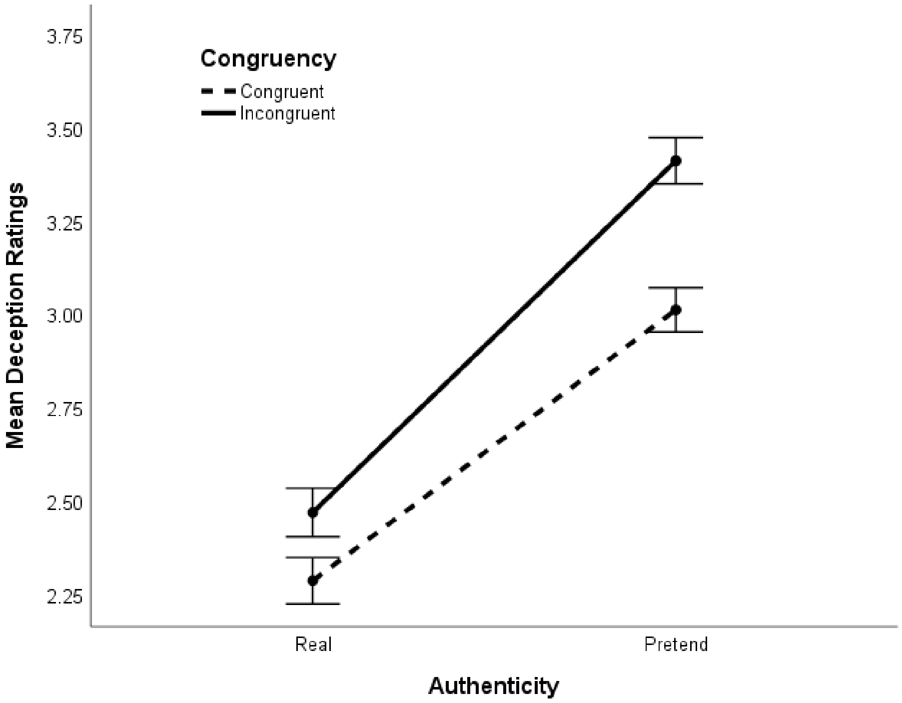
Figure C.2

Interactions between Authenticity and Congruency for (a) ASC and (b) NT receivers

(a)



(b)



Note. The scale of the graphs has been adjusted to illustrate the differences between categories better; the Y-axis does not start at zero. Both (a) and (b) are on the same scale. Error bars represent one standard error.

C.4. Investigating induced intergroup bias: WP-DEP and TTL-DEP

As the data from the WP-DEP and the TTL-DEP was ordinal, a 2 (congruency) x 2 (receiver diagnosis) ANOVA was not possible. Therefore, the same analyses as the original data was conducted here: a between subjects Mann-Whitney U-test was employed to test the main effect of receiver group, and a within subjects Wilcoxon signed-rank test was used to test the main effect of congruency. To test the interaction between the two factors, a between subjects Mann-Whitney U between ASC and NT receivers on the difference between their congruent and incongruent score. To avoid type 1 errors due to the increased number of comparisons, a more stringent alpha value of 0.01 was used to ascertain significance.

C.4.1. Manipulation checks

To check if the manipulation of the “fact files” was successful, participants ratings on how autistic they thought each sender was of analysed for autistic-labelled (ASC-L) and non-autistic labelled (NT-L) senders. Two repeated measures *t*-tests indicated that the manipulation was successful in both tasks. In the WP-DEP, participants rated ASC-L senders (M = 53.17, SD = 18.72) significantly higher on autistic appearance versus NT-L senders (M = 36.71, SD = 18.48); t (174) = 7.84 $p < .001$, d = .593. Similarly, in TTL-DEP too ASC-L senders (M = 55.00, SD = 20.08) were rated higher on autistic appearance than NT-L senders (M = 40.27, SD = 18.35) as well; t (174) = 7.63 $p < .001$, d = .577.

C.4.2. Performance against chance

For the task to be deemed to be working, non-autistic (NT) participants would need to score above chance. By design, NT receivers expected to perform worse in the incongruent condition i.e. for ASC-L senders. Therefore, only the NT receivers’ performance on the NT-L condition were tested against chance to gauge if the tests were working (Table C.4).

Table C.4

Non-autistic receivers' performance against chance on congruent condition on novel tasks

	n	*M* (*SD*)	*Mdn*	*T*	*z*	*p*	*r*
WP-DEP	97	0.87 (0.66)	1	2835.00	1.69	.091	0.17
TTL-DEP	97	1.05 (0.76)	1	3378.00	3.66	<.001	0.37

Note. WP-DEP = Weird Pictures (Double Empathy Problem), TTL-DEP = Two Truths and a Lie (Double Empathy Problem).

One-sample Wilcoxon Signed Rank tests were performed against a test value of 0.67 (chance: 2 trials x 1/3 probability of getting the correct answer by chance). NT receivers scored significantly above chance for the TTL-DEP, but their performance was not significantly above chance for WP-DEP.

Considering that NT receivers were unable to perform above chance on the WP-DEP, the findings from WP-DEP were considered to be uninterpretable as the task was deemed to be not working. Therefore, no further analyses on the WP-DEP were conducted, and henceforth only the data collected on the TTL-DEP was analysed.

C.4.3. Main effect of receiver diagnosis

A Mann-Whitney U-test was run to check if the receiver groups differed on how well they judged lies, regardless of the sender labels. Participants could obtain a score between 0 to 4. Autistic receivers (M = 1.77, SD = 1.04, Mdn = 2) showed a tendency to detect lies slightly less often than non-autistic receivers (M = 2.09, SD = 1.12, Mdn = 2). However, this difference was not significant, U = 4332.00, z = 1.71, p = .087, r = .13.

C.4.4. Main effect of congruency

A Wilcoxon signed-rank test was run to see if lie detection differed depending on congruency between sender label and receiver diagnosis. As there were only 2 NT-L trials and 2 ASC-L trials, participants could achieve a score between 0-2 for each of the congruency conditions.

Receiver's lie detection in TTL-DEP was not significantly affected by congruency, i.e. no difference in lie detection between the congruent condition (M = 0.92, SD = 0.77, Mdn = 1) and the incongruent condition (M = 1.03, SD = 0.72, Mdn = 1), T = 3243.00, z = 1.36, p = .173, r = .10.

C.4.5. Interaction between receiver diagnosis and congruency

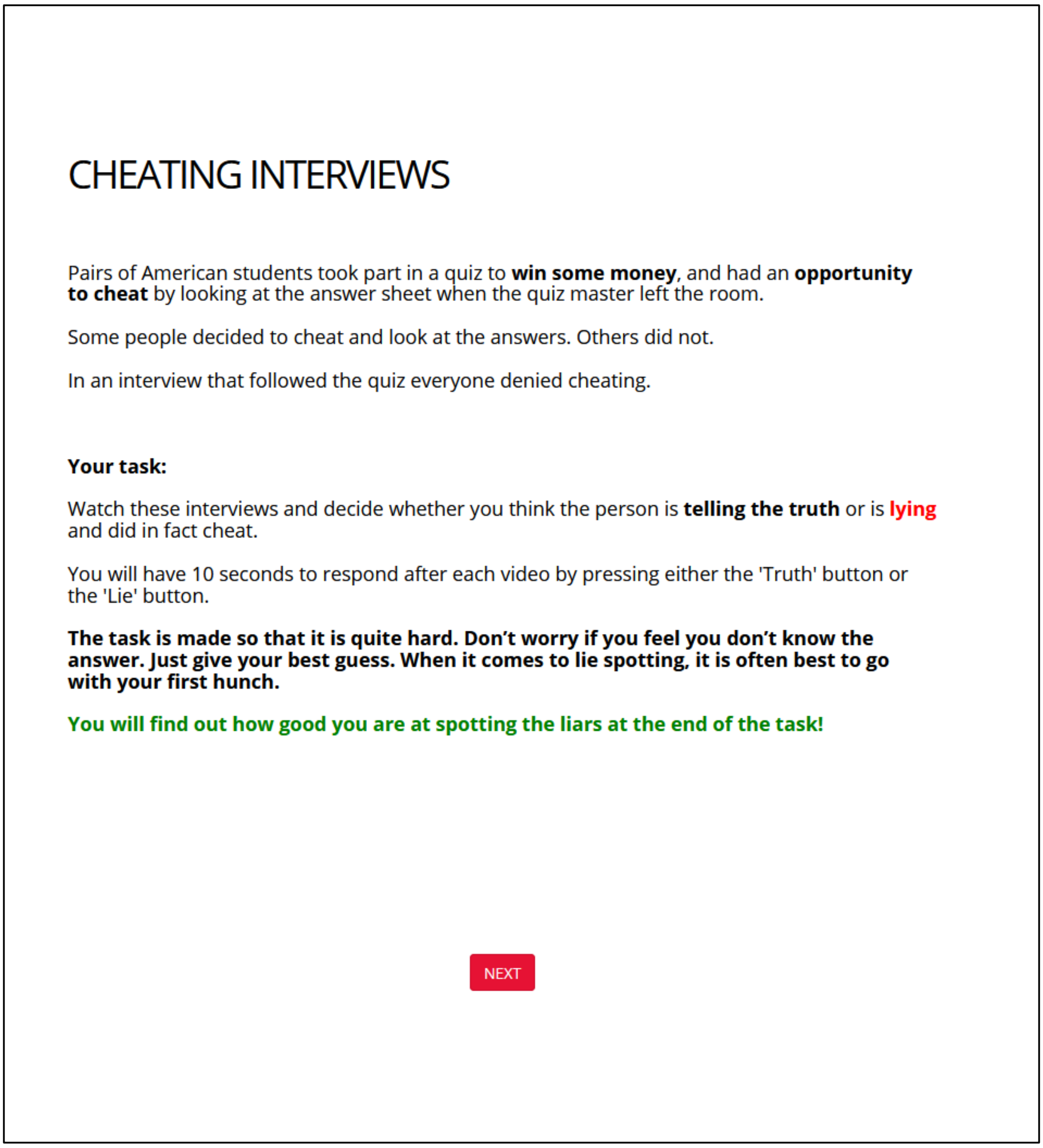
To examine if congruency between sender labels and receiver diagnosis affected performance differently for autistic and non-autistic receivers, an 'interaction score' was calculated by subtracting the receivers' scores in incongruent conditions from their scores in congruent condition. This score could range from -2 to +2. A higher score denoted better lie detection for congruent senders. A significant difference between groups in their interaction scores would indicate that congruency affected lie detection performance differently in each group.

A Mann-Whitney U-test demonstrated that autistic receivers (M = -0.26, SD = 1.09, Mdn = 0) and non-autistic receivers (M = 0.01, SD = 0.94, Mdn = 0) did not significantly differ in how congruency affected their performance in the TTL-DEP, U = 4282, z = 1.57, p = .117, r = .12. These results must be interpreted in the context of the main effect of congruency. As there was no effect of congruency – i.e. across groups, receivers did not show any difference in their deception detection abilities in either congruent or congruent conditions – the non-significant result from the interaction scores demonstrates that neither group performed better in any condition. In other words, the results from the TTL-DEP show that both groups performed comparably in all conditions, replicating the results from the original analyses.

Appendix D: Deception Task Instructions

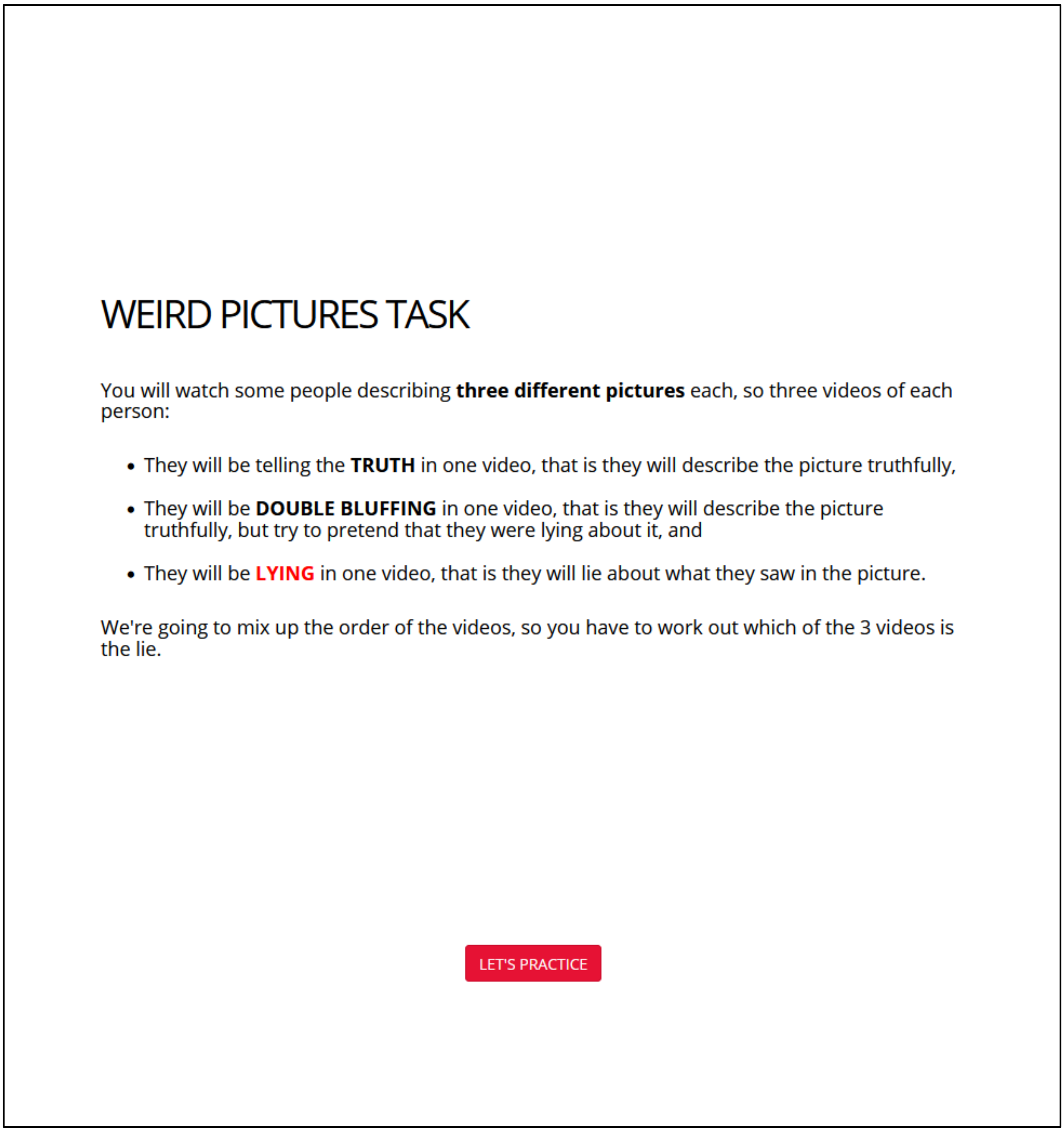
D.1. Cheating Task

Participants watched a video with detailed instructions on how to attempt the tasks. Below is a screen shot of the screen with the short written instructions that was displayed after the screen with the instruction video.



D.2. Weird Pictures (Development of Deception Detection)

Participants watched a video with detailed instructions on how to attempt the tasks, which also included an example ‘Weird Picture’ that the senders described. Below is a screen shot of the screen with the short written instructions that was displayed after the screen with the instruction video.



D.3. Weird Pictures (Double Empathy Problem)

The WP-DEP did not have video instructions as they were only administered to adults, so detailed written instructions were provided and reiterated again before practice trials and experimental trails. Below are the screenshots of these screens.

WEIRD PICTURES TASK

Each person sat in front of another person and described different weird pictures to their partners. They were required to describe every picture for approximately 30 seconds. A bell told them when to start describing and when to stop.

There were three ways they could describe every picture. They could either tell:

- the **TRUTH**: They described the picture truthfully.
- a **DOUBLE BLUFF**: They described the picture truthfully, but tried to pretend like they were lying about it.
- a **LIE**: They described the picture untruthfully, that is, they lied about what they saw in the picture.

NEXT

Practice Trial

You will be shown a 'Fact file' about 1 person, followed by a video where they describe three 'weird pictures'. You have to guess which of the three descriptions is the lie. You will be directed to the response screen automatically once the video finishes playing. You will have **5 seconds** to respond.

You will also be asked some general questions about your impression of the person in the video.

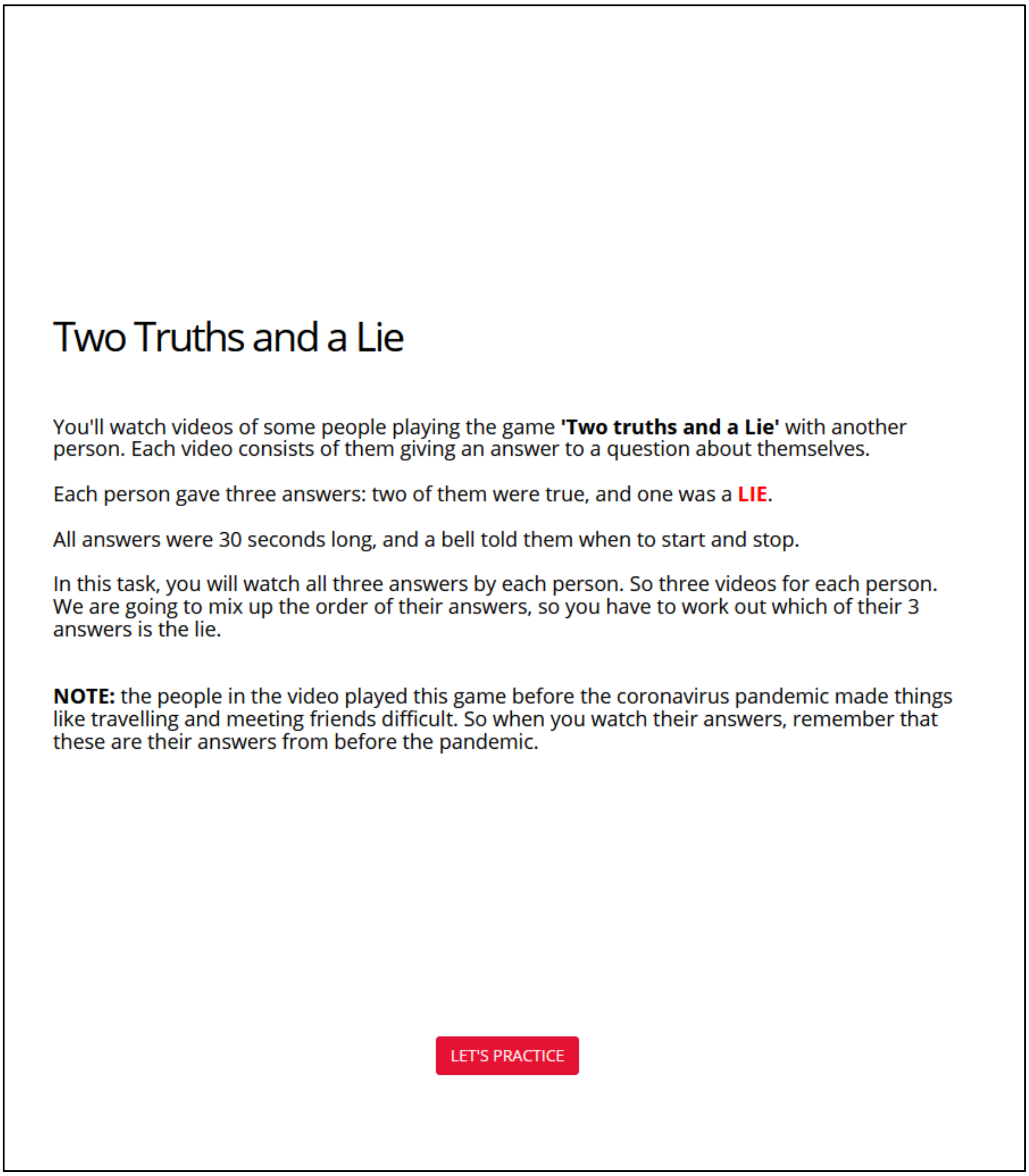
Please make sure you have your volume turned up and, if using earphones, you are wearing them.

Click START to begin the practice trial.

START

D.4. Two Truths and a Lie (Development of Deception Detection)

Participants watched a video with detailed instructions on how to attempt the tasks. Below is a screen shot of the screen with the short written instructions that was displayed after the screen with the instruction video.



Two Truths and a Lie

You'll watch videos of some people playing the game **'Two truths and a Lie'** with another person. Each video consists of them giving an answer to a question about themselves.

Each person gave three answers: two of them were true, and one was a **LIE**.

All answers were 30 seconds long, and a bell told them when to start and stop.

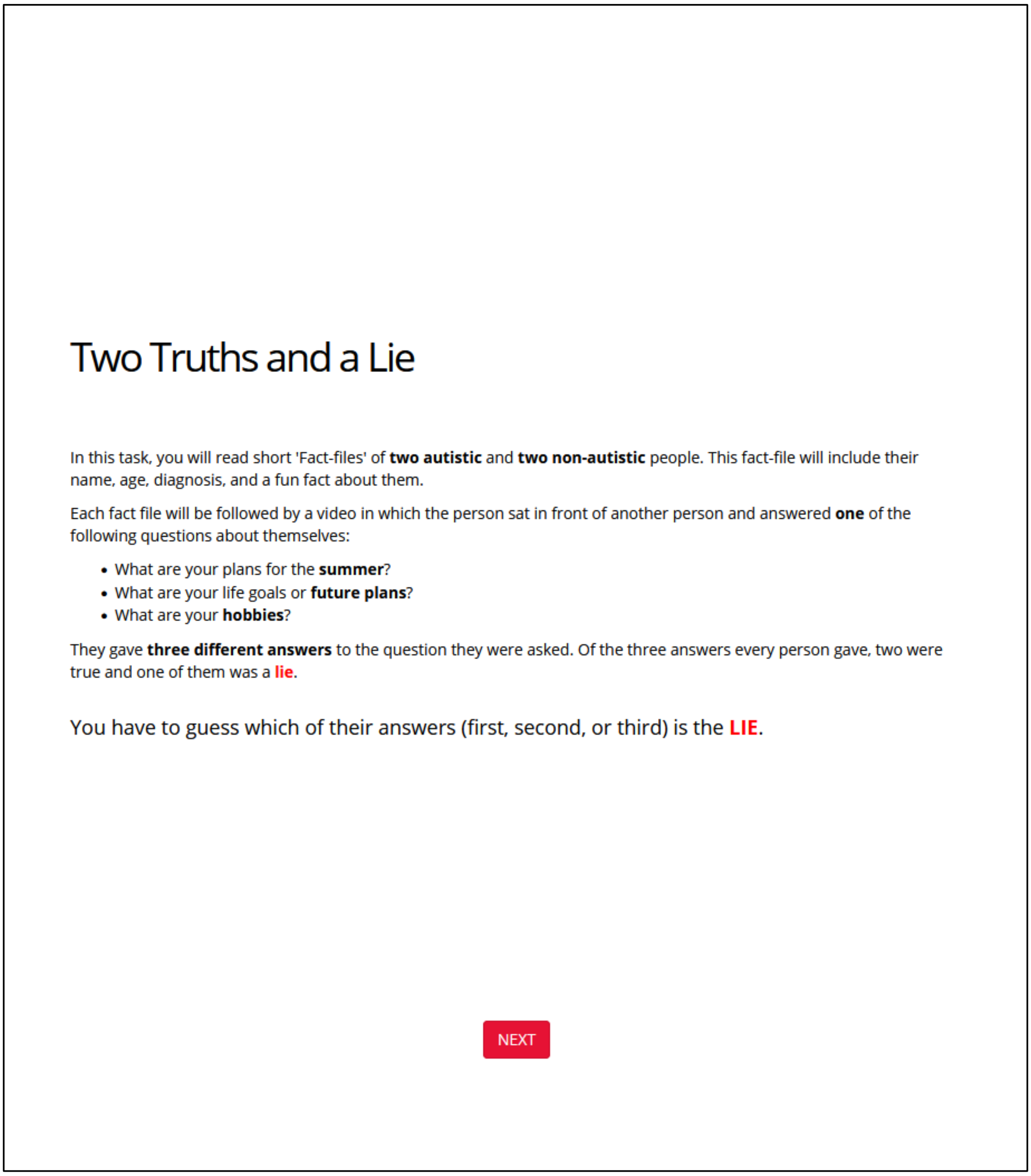
In this task, you will watch all three answers by each person. So three videos for each person. We are going to mix up the order of their answers, so you have to work out which of their 3 answers is the lie.

NOTE: the people in the video played this game before the coronavirus pandemic made things like travelling and meeting friends difficult. So when you watch their answers, remember that these are their answers from before the pandemic.

LET'S PRACTICE

D.5. Two Truths and a Lie (Double Empathy Problem)

The TTL-DEP did not have video instructions as they were only administered to adults, so detailed written instructions were provided and reiterated again before practice trials and experimental trails. Below are the screenshots of these screens.



Two Truths and a Lie

Each of the three answers a participant provided is approximately 30 seconds long. While watching the videos, you might hear a bell at the beginning or end of a video - this bell told the participant when to start or end an answer.

NOTE: These videos were shot before the 2020 Coronavirus pandemic, and therefore the answers given by the people in the video do not reflect the changes brought on by the pandemic.

Click **NEXT** to go through a practice trial and watch an example video.

NEXT

Practice Trial

You will be shown a 'Fact file' about 1 person, followed by a video where they answer a question with three different answers. You have to guess which of their answers (first, second, or third) is the **LIE**.

You will be directed to the response screen automatically once the video finishes playing. You will have **5 seconds** after the video finishes playing to give your response by clicking one of three buttons.

You will also be asked some general questions about your impression of the person in the video.

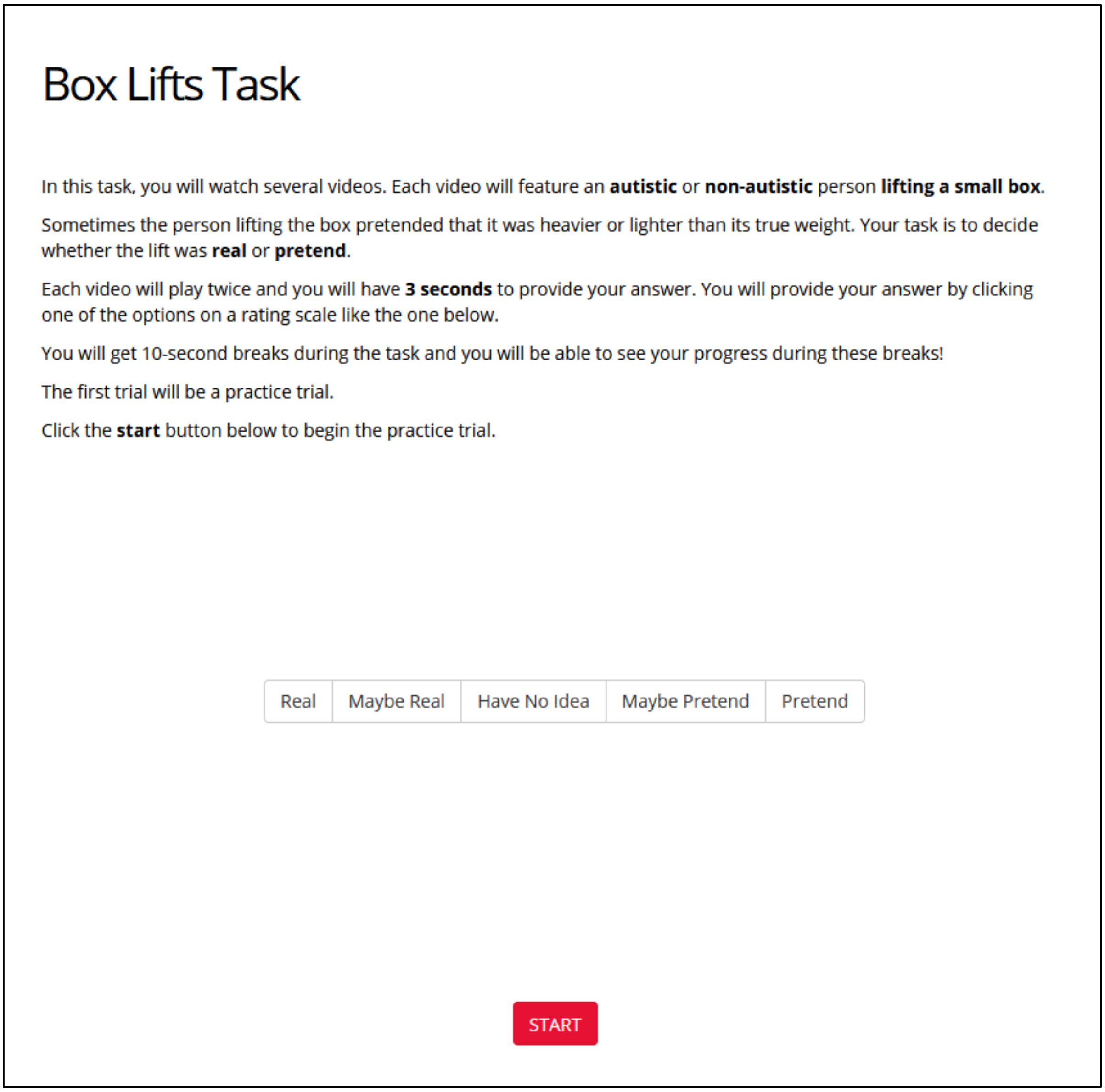
Please make sure you have your volume turned up and, if using earphones, you are wearing them.

Click START to begin the practice trial.

START

D.6. Box Lifts (Double Empathy Problem)

The Box Lifts tasks did not have video instructions as they were only administered to adults, so detailed written instructions were provided at the start of the task. Below is a screenshots the instruction screen.



Box Lifts Task

In this task, you will watch several videos. Each video will feature an **autistic** or **non-autistic** person **lifting a small box**.

Sometimes the person lifting the box pretended that it was heavier or lighter than its true weight. Your task is to decide whether the lift was **real** or **pretend**.

Each video will play twice and you will have **3 seconds** to provide your answer. You will provide your answer by clicking one of the options on a rating scale like the one below.

You will get 10-second breaks during the task and you will be able to see your progress during these breaks!

The first trial will be a practice trial.

Click the **start** button below to begin the practice trial.

Real	Maybe Real	Have No Idea	Maybe Pretend	Pretend

START

Appendix E: Resource Sheets

As this was an online study, there was an ethical concern regarding the participants' wellbeing after answering difficult questions about peer-victimization and mental health. Resource sheets were provided to adult and adolescent participants, as well as parents of adolescent participants. These were presented as part of the task as well as pdf versions that were available for participants to download. These documents are presented below.

INSTITUTE OF COGNITIVE NEUROSCIENCE

Resource Sheet for Parents of Participants under the age of 18 years

Dear Parent,

Thank you for showing interest in our project “Spotting Lies and Reading Minds”, and for considering volunteering your child to participate in our study. By completing these tasks, your child would help answer some research questions about:

- How different people interpret different social scenarios
- How different people understand what another person is thinking
- How effective people are at detecting lies in others
- How these abilities change throughout development
- How these abilities differ in people with and without autism

As mentioned in the Information Sheet, there might be certain questions asked during the study which might be sensitive in nature. The questionnaires will touch upon issues relating to bullying, vulnerability, victimization and mental health. As such, there is a chance that your child might become upset on having to answer these questions if they identify themselves as someone who has been in a situation that constitutes as bullying, have felt vulnerable to harassment or experience mental health related difficulties. We will encourage your child to come and speak to you or another trusted adult should they feel upset by any of the activities or questions.

Talking to your children about their experience

We would encourage you to engage with your child when they have finished the activities to ensure that they are not in any way distressed, or give them an opportunity to speak to you regarding any concerns they might have. Here are some resources you can look at to help you talk to your child about bullying and mental health should you wish to:

Bullying.co.uk: how to talk to your child about bullying

Kidscape.org.uk: talking about bullying with your child

Actionforchildren.org.uk: how can you help with children and young people’s mental health

Huffingtonpost.co.uk: speak to your kids about their mental health

HELPLINES:

If relevant, we also recommend reaching out to charities and societies who can offer additional help and advice. We will similarly provide your child with links to resources about mental health and bullying, and some charities and societies they can contact for more information after they take part in the study. You may find the following to be a valuable resource:

A charity which runs an out-of-hours helpline offering specialist emotional support and information to anyone affected by mental illness.
Telephone helpline: 0845 767 8000 (open 6pm-11pm, everyday)
Website: http://www.sane.org.uk/

Advice for parents and carers worried about a child or young person under 25, regarding a child's behaviour, emotional wellbeing, or mental health condition.
Parent's telephone helpline: 0808 802 5544
Website: http://www.mind.org.uk/

Helpline open 24 hours a day, 365 days of the year.
Telephone Helpline: 020 8394 8300
Website: http://www.samaritans.org

A charity providing support to both adults and children with bullying at home, in the community, the playground, the workplace or on-line.
Telephone helpline: 0845 22 55 787 (Open 9am to 5pm Monday to Friday)
Website: https://www.nationalbullyinghelpline.co.uk/

Advice for school, workplace and cyberbullying.
Telephone helpline: 0808 800 222
Website: https://www.bullying.co.uk/

A charity providing support to people with autism
Website: http://www.autism.org.uk/

INSTITUTE OF COGNITIVE NEUROSCIENCE

UCL

Debrief Sheet for Participants under the age of 18 years

Dear Participant,

Thank you for taking part in our project "Spotting Lies and Reading Minds". We hoped you enjoyed the tasks. By completing this study, you helped answer some research questions about:

- How different people interpret different social situations
- How different people understand what another person is thinking
- How affective people are at catching when others are lying
- How these abilities change as and when people grow up
- How these abilities may be different in people with and without autism

We would love to know on how you found the activities.

- Did the study run smoothly? Did something not work?
- Was it clear what you were supposed to do in the activities and were the tasks doable?
- Did you find any activities stressful or too difficult?
- Did any of the topics discussed or the questions asked make you feel upset?

What should you do if some part of this study made you feel upset?

If any of the topics discussed or any of the questions asked **made you to feel upset, sad or distressed, please talk to a parent or another trusted adult**. They would most likely ask you about how you found the tasks and about your experience in participating in the study, and this might be a good time to discuss with them if you are upset about the questions you were asked and how the topics made you feel.

If you would like to get in touch with one of the researchers regarding this, just ask your parents to email us, and we would be happy to help you.

HELPLINES:

If you would rather talk to someone about these feeling who is not a parent or another adult close to you, there are various charities and groups of people who **offer help and advice to children and young people about different issues, including mental health and bullying**. You can go on to their website to read what they have to say or call them up. Here are some of them:

YOUNGMINDS fighting for young people's mental health

A charity providing information about getting help with mental health problems and accessing services.
Website: http://www.mind.org.uk/

childline ONLINE, ON THE PHONE, ANYTIME

A free, private and confidential service where anyone under the age of 19 can talk about anything.
Telephone Helpline: 0800 1111
Website: https://www.childline.org.uk/

NATIONAL BULLYING HELPLINE

A charity providing support with bullying at home, in the community, the playground, the workplace or on-line.
Telephone helpline: 0845 22 55 787 (Open 9am to 5pm Monday to Friday)
Website: https://www.nationalbullyinghelpline.co.uk/

BullyingUK Part of family lives

Advice for school, workplace and cyberbullying.
Telephone helpline: 0808 800 222
Website: https://www.bullying.co.uk/

National Autistic Society

A charity providing support to people with autism
Website: http://www.autism.org.uk/

INSTITUTE OF COGNITIVE NEUROSCIENCE

Debrief Sheet for Adult Participants in Research Studies

Principal Investigator: Dr. Sarah White, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR. Email: s.white@ucl.ac.uk. Tel 020 7679 1148

Researchers: Ms. Ishita Chowdhury and Ms. Malwina Dziwisz, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR. Email: ishita.chowdhury.16@ucl.ac.uk, m.dziwisz@ucl.ac.uk

UCL Data Protection Officer: data-protection@ucl.ac.uk

This project has been approved by the UCL Research Ethics Committee as Project ID 14807/002

Dear Participant,

Thank you for taking part in our project “Spotting Lies and Reading Minds”. By completing these tasks, you helped answer some research questions about:

- How different people interpret different social scenarios
- How different people understand what another person is thinking
- How effective people are at detecting lies in others
- How these abilities change throughout development
- How these abilities differ in people with and without autism

We would love to have some feedback on how you found the day’s activities.

– Did the study run smoothly? Did something not work?
– Were the activities presented clearly and were the tasks achievable?
– Did you find any activities stressful or too difficult?
– Did any of the topics discussed or the questionnaires make you feel upset?

What should you do if the task in this study caused you distress?

If any of the topics discussed or questionnaires made you feel upset or caused you any distress, we would strongly recommend speaking with a trusted friend, family member or colleague. If you would like to get in touch with one of the researchers regarding this, please email us, and we would be happy to help you.

If you are worried about your mental health, the best option would be to see your GP and discuss with them how you are feeling. They will be able to assess how you are feeling and discuss the potential options that are available to you.

it is possible to directly access mental health support in the UK through NHS IAPT: https://www.nhs.uk/service-search/find-a-psychological-therapies-service/

HELPLINES:

If relevant, we also recommend reaching out to charities and societies who can offer additional help and advice. You may find the following to be a valuable source:

SANE

A charity which runs an out-of-hours helpline offering specialist emotional support and information to anyone affected by mental illness.
Telephone helpline: 0845 767 8000 (open 6pm-11pm, everyday)
Website: http://www.sane.org.uk/

mind
for better mental health

A charity providing information about getting help with mental health problems and accessing services.
Website: http://www.mind.org.uk/

SAMARITANS

Helpline open 24 hours a day, 365 days of the year.
Telephone Helpline: 020 8394 8300
Website: http://www.samaritans.org

NATIONAL BULLYING HELPLINE

A charity providing support to both adults and children with bullying at home, in the community, the playground, the workplace or on-line.
Telephone helpline: 0845 22 55 787 (Open 9am to 5pm Monday to Friday)
Website: https://www.nationalbullyinghelpline.co.uk/

Advice for school, workplace and cyberbullying.
Telephone helpline: 0808 800 222
Website: https://www.bullying.co.uk/

A charity providing support to people with autism
Website: http://www.autism.org.uk/