You are Lying! Verbal Deception Detection Within the Investigative Interview: Testing the Differences in Provided Verbal Cues Between Autistic and Neurotypical Suspects

Lennart Dignath (s2743833)

Department of Behaviour, Management and Social Sciences, Section of Psychology of Conflict, Risk and Safety

Master Psychology Conflict, Risk and Safety

University of Twente

1st Supervisor: Dr Lynn Weiher

2nd Supervisor: Dr Steven Watson

06 July 2025

Abstract

In the investigative interview, one of the main goals is to gather reliable information and details. However, suspects might try to mislead the interviewer or hold back information. Therefore, trying to detect deception is a crucial aspect in investigative interviewing. A method used in research for deception detection is reality monitoring (RM), focusing on differences in the number of details to distinguish truthtellers from deceivers. The number of details someone mentions can be influenced by cognitive load, decreasing the amount of information the interviewee can hold in their working memory. Therefore, due to difficulties in cognitive load, autistic individuals might deceive differently, have difficulties with providing lies or provide fewer details due to encoding difficulties which could reduce the effectiveness of RM, since the testimony of an autistic individual might be incorrectly classified as a lie due to the vulnerability. This paper is among the first to research the use of RM with autistic suspects. We used a 2 (autistic vs. neurotypical) x 2 (guilty vs. innocent) quasi-experimental design in which students were interviewed in a mock-investigative interview. The results showed that autism did not influence the number of details provided and that cognitive load had no effect on the relationship between autism and the number of provided details. Overall, autistic individuals provided slightly more external details than non-autistic individuals, however, this effect was not significant. Moreover, condition (guilty vs. innocent) had a significant effect on the number of provided details. These findings indicate that autism seems to not influence the number of provided details. As expected, being guilty reduced the amount of provided details.

Keywords: investigative interview, deception, deception detection. ASD, neurodiverse, RM, Reality Monitoring

Introduction

We acknowledge that there are differences in referring to autistic people. Therefore, we decided to use both, person first and identity first language, in this paper.

One of the main aims of the investigative interview is to obtain a great deal of reliable details (Weiher et al., 2023). However, suspects might try to lie in the investigative interview denying or avoiding incriminating information or trying to mislead the interviewer (Vrij et al., 2014; Hartwig et al., 2014). Therefore, it is important to be able to detect deception to ensure the quality of the gathered information in the interview. A method used to detect deception in research is reality monitoring (RM) focusing on differences in memories between experienced and fabricated events (Johnson & Raye 1981). However, Norris and Maras (2022) argue that neurodivergent individuals have difficulties retrieving event-related memories, which might indicate that RM is not suitable for detecting deception in neurodivergent individuals. Moreover, due to increased cognitive load for neurodivergent individuals, lying might in general be more difficult and it has been shown that autistic individuals find lying more difficult than neurotypical individuals (Bagnall et al., 2022) The aim of this paper is to test whether RM can be used for deception detection with neurodivergent individuals since it is important how neurodivergent individuals might differ in their verbal deception to improve research and eventually help practitioners in correctly and effectively detecting deception and gathering high-quality information in the investigative interview.

Deception Detection within the Investigative Interview

Investigative interviewing is an important part of the investigation since it helps gathering information from suspects and witnesses, with a great deal of reliable details, which is one of the key purposes of the interview (Vrij et al., 2014; Weiher et al., 2023). However, the elicited details need to be checked for veracity to determine if the gathered information is reliable. Therefore, detecting if the interviewee is trying to deceive the interview is important, since suspects of crime may try to withhold information or try to mislead the interviewer which makes it important to detect whether a suspect is trying to deceive the interviewer (Vrij et al., 2014).

Suspect may use different deception strategies, such as concealing information (Hartwig et al., 2014). Often, suspects aim to appear innocent by managing the information they disclose, which creates an information management dilemma (Hartwig et al., 2014) since the suspect has more knowledge than they are willing to share. To manage this dilemma, liars may avoid revealing certain information or denying their knowledge (Hartwig et al., 2014). Knowing what methods deceivers might use to deceive can inform how we can detect this deception. When trying to detect deception, often non-verbal methods are used. We know methods such as the polygraph which try to detect deception based on the idea that there is a relationship between physiological changes which manifest when a person is lying. However, such changes in physiological state can be induced by other states than lying (e.g., fear) (Brewer & Williams, 2005), which makes these methods less reliable. In a meta-analysis, Sporer and Schwandt (2007) found that directly observable behaviour (e.g., hand or leg movements) are negatively correlated with deception which indicates that liars take part less often in these behaviours. Moreover, Luke (2019) has shown that for many of the cues estimated to at least weakly show deception it is not so certain that these cues have an effect. Due to low power, few replications and publication bias we do not know for certain that studied cues have an effect other than zero. This calls for scepticism in the research of deception cues. Therefore, we see that measures focusing on behavioural or physiological indicators for deception seem to be insufficiently justifiable.

However, methods using verbal cues to detect deception seem more promising. Interviewing methods focusing on verbal cues to detect deception such as Reality Monitoring have proven to identify deception well. Oberlader et al. (2016) performed a meta-analysis about the predictive accuracy RM, finding evidence that content-based techniques can discriminate between truthful or fabricated statements. Statement evaluation methods such as RM have provided reliable results that these methods can detect deception. Therefore, in this paper, we will focus on the use of verbal cues to detect deception since these seem to be a promising option for deception detection, more specifically on RM. We will focus on the RM since this method can be used after an interview being a statement evaluation method. Therefore, it can detect deception without the need for a specific interviewing method.

RM focuses on memories and states that the truth is a recollection of something someone has done or witnessed, and a lie is a self-made internal memory. Johnson and Raye (1981) explain that the origin (i.e. external or imagined) of a memory can be found in the features of this memory. External memories will contain more sensory, contextual or semantic details and imagined memories more mentioning of cognitive processes (i.e. mentioning of reasoning, imagination or thought processes) since the truth (i.e., external memory) is a recollection of something experienced and a lie (i.e., imagined memory) is an internally generated memory (Masip et al., 2005). However, according to Masip et al. (2005) there is no standard set of criteria yet for this method and RM seems to be mostly used in research and not in practice (Vrij et al., 2018). There is also no standard cut off score for a specific number of details to determine if someone is deceiving or not (Oberlader et al., 2016). Moreover, there still appears to be a relatively high risk of mis-classifying witnesses since too many variables (e.g., delay or opportunity to prepare an account) can influence the accuracy of RM and not all criteria of RM are highly distinctive between deception and truth telling (Masip et al., 2005). Therefore, systematic research is still needed.

On the other hand, there is evidence that RM can discriminate deception above chance level Masip et al., (2005) performed a meta-analysis focusing on the RM approach, stating that the overall RM system discriminates well between truthful and deceiving participants. Some RM criteria seemed to be better in discriminating (e.g., contextual or spatial information) and others are less reliable (e.g., cognitive operations). Moreover, Oberlader et al., (2016) performed a meta-analysis on RM in which they focused on the final judgment about credibility. This resulted in an overall effect size of $g = 1.03$ which is large and found to be significant, indicating that when using RM deception was detected more successfully than without or using other models. When assumed that sensitivity and specificity are equal this would result in correctly identifying 70% of lies and truths (Oberlader et al., 2016). Therefore, we might say that verbal approaches such as the RM seem to be useful to further research and promising in the field of deception detection, since the RM has been shown to be useful with neurotypical individuals, however, vulnerable people like autistic individuals are seven times more likely to encounter with the criminal justice system (CJS) (Debbaudt, 2002; Brewer & Young, 2015).

We do not know if the RM can accurately detect deception for vulnerable individuals from the current research.

Vulnerability

Gudjonsson (2006) defines vulnerability as follows: “psychological characteristics or mental state which render a witness prone, in certain circumstances, to providing information which is inaccurate, unreliable or misleading” (p. 68). However, this also is applicable to suspects of a crime. Suspect with vulnerabilities might not be able to cope well with the stress in the investigative interview or feel highly anxious. Vulnerable individuals might also struggle more with disapproval, anger or threats made by the interviewer, which becomes more likely when the interviewer might become frustrated by poor performance of the interviewee (Herrington & Roberts, 2012; Gudjonsson, 2003). This can lead to increased uncertainty of the vulnerable suspect which can result in contradiction, suggestibility or even false confessions (Herrington & Roberts, 2012).

One type of vulnerability is Autism Spectrum Disorder (ASD). The National Institute of Mental Health (2024) states that ASD is a neurological and developmental disorder which affects how individuals interact, communicate, learn and behave. As adolescents and young adults, ASD individuals might have difficulties in maintaining friendships or communicating with peers and adults. ASD individuals might suffer from other conditions such as anxiety or attention-deficit/hyperactivity disorder (ADHD), occurring more often occur in individuals with ASD (U.S. Centers for Disease Control and Prevention, n.d.). According to Lim et al. (2022), neurodivergent individuals are often judged as guilty or deceiving. It might be that there is something in the presentation of autistic individuals which contradicts the observer’s perception of how a truthful individual should behave (Lim et al., 2022). Naturalistic behaviours for ASD individuals such as gaze-aversion or fidgeting might be misinterpreted by neurotypical individuals as signs of deception (Logos et al., 2021) Moreover, ASD individuals are at a higher risk to get into contact with the CJS and then being perceived as deceptive, while often finding it more difficult to deceive than neurotypical individuals (Debbaudt, 2002; Brewer & Young, 2015; Lim et al., 2022; Bagnall et al., 2022). Therefore, in this paper, we will focus on ASD.

RM focuses on the recall of details based on the memory of a specific event to identify cues for deception. However, ASD individuals can have difficulties with the recall of episodic memories (i.e. memories of specific events) which are probably related to problems with retrieving information. When recalling episodic memories autistic adults seem to recall less exact memories with fewer detail (Norris & Maras, 2022). Coutelle et al. (2021) conducted a study in which autistic participants and non-autistic participants had to recall six autobiographic memories spontaneously, using an autobiographical memory task. The participants were then cued using specific cueing which is the use of specific questions retrieving details related to predefined strategies to elicit details. The results showed that when not cued there was no difference in number or richness of details between ASD and the control group. However, when a specific cueing procedure was used the internal details were lower in the ASD group than the control group, since less contextual details were mentioned. These memory problems seem to be most prevalent in situations with unsupportive questioning which do not help in retrieving memory (e.g., investigative interviewing which uses open questions, such as “tell me everything that happened”) (Norris & Maras, 2022). Hence, the type of questioning might already influence the number of provided details. In addition, ASD individuals have more difficulties in recalling episodic memory and recall less details. Therefore, it can be argued that the applicability of the RM framework for neurodivergent suspects could be impaired, since the differences in details might not be a valid distinguishment between truth telling and lying for ASD individuals. Therefore, RM might falsely classify truth telling ASD individuals for being deceptive.

Moreover, the investigative interview might be more difficult regarding communication and interaction for ASD individuals than for neurotypical individuals. Therefore, ASD suspects might naturally experience difficulties with cognitive processes and find lying even more demanding, struggling with inhibiting truthful responses or holding lies in short-term memory (Blackhurst et al., 2024) Therefore, ASD suspects might be less able to tell lies in the investigative interview. Moreover, cognitive control for ASD individuals might be lower which refers to mental operations prioritising information which is essential for adaptive and flexible behaviour in uncertain situations. In the study performed by Mackie & Fan (2016) it was shown that ASD individuals performed less efficient and had less cognitive control, which could affect the information management dilemma (i.e., what kind of

information does the suspect avoid and/or denies). Therefore, the ASD suspect might be less able to respond more flexible to new evidence disclosure or come up with new information due to the decreased cognitive control. This might affect the different kind of details and number of details provided in the statement of the suspect, which again could influence the accurateness of the RM method.

This research paper will therefore try to answer the following research question: *"Can Reality Monitory accurately detect deception in individuals with Autism Spectrum Disorders, considering the impairments in recall of episodic memory and cognitive load?"*. This research if one of the first to study deception detection with ASD individuals. The research question translates into the following hypotheses which will be tested. ASD individuals have impaired retrieval of episodic memory and unsupportive questioning increases the difficulties to retrieve episodic memory. Therefore, the number of details about an event might be reduced in accounts provided by ASD individuals. Thus, we predict that the accounts of ASD individuals in general will have less contextual details than the accounts of individuals without ASD. The main hypothesis of this paper is that:

H1: ASD Individuals will provide significantly less contextual details than neurotypical individuals regardless of the veracity status (truthful or deceptive).

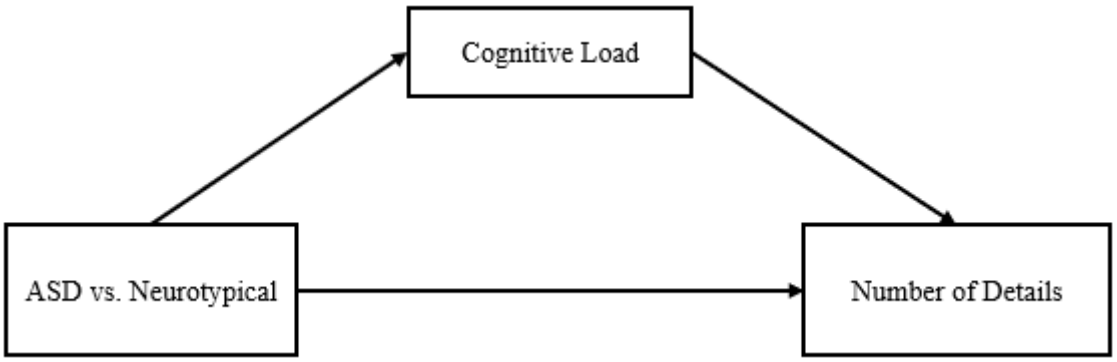
H1a: ASD Individuals will report significantly higher cognitive load during the investigative interview than neurotypical individuals, which will predict fewer overall details in their accounts when deceiving.

Since cognitive load is known to impair memory functions, we predict that cognitive load will act as a mediator in the relationship between ASD and the number of details (See Figure 1), impacting the recall of details. We predict that individuals with ASD will experience a heightened cognitive load due to their vulnerability being less able to cope with the stress in the investigative interview. This cognitive load will affect the number of details in the account.

H1b: The difference between contextual and internal details in deceptive accounts will be significantly smaller for ASD individuals than for neurotypical individuals, indicating a reduced effectiveness of RM indicators for deception detection in the ASD group.

Figure 1

The Relationship Between ASD and Number of Details with Cognitive Load as a Mediator



We predict that providing less contextual details will lead to smaller differences between contextual details and internal/cognitive details. Therefore, the difference between internal and contextual details in the accounts of deceiving individuals with ASD will be too small to detect this deception. This is an important indication in the RM method, indicating that the usefulness of RM might be lower for ASD individuals.

Method

Design

This study used a 2 (Neurotype: autistic vs. non-autistic) x 2 (Condition: guilty vs. innocent) mixed methods between-subjects quasi-experimental design with as dependent variables internal and external memories and was conducted by two experimenters. This is a quasi-experimental design since the group neurotype cannot be randomized. The research got ethical approval from the University of Twente (request – 250097 and 240950)

Participants

The participants were sampled using convenience, snowball and voluntary sampling via referring of peers, flyer advertising, , WhatsApp groups and the SONA system of the University of Twente. Participants taking part via this system got granted course credits. Other than this no compensation was used. The inclusion criteria a minimum age of 18 years to provide informed consent and participants should have adequate English language skills.

Audio data for five participants was not available due to data corruption or recording issues. After removing, the sample consisted out of 97 participants, 68 (70%) participants had a Dutch nationality, 11 (11%) participants German and 18 (19%) participants indicated to have another nationality, mostly being European, but also participants from Brazil, Guatemala or Tunesia. Forty-seven (48.5%) participants indicated to be female and 50 (51.5%) indicated to be male. The participants had a mean age of 23.3 (*SD* = 2.50). Forty-eight participants took part in the guilty condition and 49 participants took part in the innocent condition. The condition (guilty vs. innocent) was allocated using block randomization. The first 34 participants were allocated into the guilty condition. Then participants were allocated randomly with a slight focus on the innocent condition to get the condition groups equal for a better comparison. In each condition, the participants were provided with a case providing a background story to assign guiltiness or innocence, which will be more explained in the materials section. Fifty were autistic, 47 were neurotypical. 19 participants with a professional diagnosis, some with multiple diagnoses were counted into the autistic conditions with allowing anxiety or AD(H)D diagnoses since these occur more often in individuals with ASD (U.S. Centers for Disease Control and Prevention, n.d.). Eight participants reported no diagnosis but suspected to be neurodiverse, and 69 participants indicated to have no diagnosis. Participants without an official diagnosis were screened using the Autism Quotient with a cut off score of six or higher (AQ-10; Baron-Cohen et al., 2001) and ADHD self-report scale screener with a cut off score of 4 or higher (ASRS v1.1; Kessler et al., 2005). Based on the screeners, including the 19 participants with an official diagnosis, there was a total of 50 autistic participants.

Materials

To simulate a realistic investigative setting, participants took part in a mock burglary scenario in a VR environment before being interviewed using a roleplay mock investigative interview.

Pre-Interview Task

We used a virtual environment which the participants could experience via a Meta Quest 2, using two handheld controllers to navigate freely in the environment. The virtual environment consisted of a neighbourhood and a house which was accessible. It was made to look like a suburban

Dutch neighbourhood in Unity 2021.3.32f. The neighbourhood consisted out of a single street, trees, parked cars and roadblocks at the end. One house could be accessed. Innocent participants were tasked to view the house as a possible house they want to buy and to look for a realtor. Guilty participants were tasked to perform a burglary in the house. To ensure that participants in the guilty condition were able to steal items in the environment, three training items (i.e. two chandeliers and one tv) were put on the street with which the guilty participants could train to grab items and put them into a white box which indicated that the participant had stolen the item. For an overview of the environment, see Appendix B.

Interview

The mock investigative interview was conducted by a single interviewer. In total, two rooms were used switching the rooms depending upon availability. The rooms were equipped with chairs, a table and a GoPro 9 to record the interview which all was made available by the BMS Lab of the University of Twente. Two rooms were used due to more availability for the experiment. To minimize the effect of potential biases, interviewer effects and familiarity with the interviewer, the two experimenters switched in their roles either leading the experiment or interviewing the participants. It was ensured that the interview did not know the participant. The interviewer used a structured interview script (see Appendix D) to ensure that interviews are conducted in the same manner for both conditions and by both interviewers. This script was developed based on the investigative interview method used in the Netherlands (Hoekendijk & Beek, 2015; Rispens & van der Sleen, 2022), following a three-step structure (i.e., (1) opening the interview providing statements as “for your protection and for mine, we will record this, so we have full account of what was said today”, (2) person-oriented interview, using questions like “do you have a job?”, and (3) case-oriented interview, asking the participant to “please tell me in as much detail as possible, why you were in the neighbourhood”. This method uses a funnel structure going from broad (i.e. getting to know the suspect and building rapport) to specific (i.e., asking more case-related questions).

Post-Interview

Screening for Neurodiversity Two self-report screeners were used to screen for ASD and AD(H)D. The Autism Quotient questionnaire consists out of 10 items using a 4-point Likert scale with items like “I often notice small sounds when others do not”, ranging from *Definitely Agree* to *Definitely Disagree*. When scoring *Definitely* or *Slightly Agree* participants were scored one point for the items one, seven and eight and when scoring *Definitely* or S*lightly Disagree* on the other items the participant scored one point. Per item one point could be scored. The cut off score to be scored with ASD was six or higher. This screener was used since it derives from the 50-item Autism Spectrum Quotient designed for measuring ASD in adults without learning disabilities and at a cut point of six the Autism Quotient questionnaire had a sensitivity of 0.88, specificity of 0.91 and positive predictive value of 0.85. This indicated that the questionnaire has the potential to aid referral decision making for professional assessment. (Allison et al., 2012).

Screening for ADHD was done using the ADHD self-report scale screener. This questionnaire consists of 6 items using a 5-point Likert scale, like “How often do you have problems remembering appointments or obligations?”, ranging from *Never* to *Always*. When scoring *Sometimes, Often* or *Always* participants were scored one point for the items one, two and three. When participants scored *Often or Always* on the other items the participant was scored one point for the item. Maximal one point per item could be scored. The cut off score to be scored with ADHD was four or higher. This screener was used since the items are derived from the 18 DSM-IV criteria for screening ADHD and these six items were seen as most predictive for symptoms consistent with ADHD (Kessler et al., 2005; Adler et al., n.d.). See Appendix E for both screeners.

Cognitive Load The cognitive load was measured using a scale created by Herrema (2015), based on the studies of Shenhav et al. (2017) and Vogels et al. (2014) measuring the ability of the suspect to verbalise thought processes and memory retrieval during the investigative interview. This scale consists of 11 items using a 5-point Likert scale, like “I found it difficult to explain the order of events while being interviewed”, ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree.* The scale has a Cronbach’s Alpha of .82 indicating high internal consistency. See Appendix F for the questionnaire.

Rapport The rapport was assessed using the Rapport scales for Investigative Interviews and Interrogations (RS3i) (Duke et al., 2018) and an interaction scale (Bernieri, 1988). The RS3i consists of 21 items on a 5-point Likert Scale, like "I think the police officer was generally honest with me", ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree*. To fit the current study the items were slightly modified (e.g., *interviewer* was changed to *police officer*). The interaction scale consists of 7 items using a 5-point Likert scale, like "The interaction seemed well-coordinated", ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree*. The data from these scales was analysed by a colleague (see Bouma, 2025) and was not included in this paper. See Appendix G for the questionnaires.

Coding Reality Monitoring The audio data of the videos was transcribed using the automatic translation service of Amberscript and afterwards controlled for mistakes from the automated service and corrected, when necessary, by hand. Potential changes were (spelling) mistakes in the transcription itself (e.g., wrong words). These transcripts were then coded using Atlas.ti version 25.0.1.32924. The coding of details was done by one experimenter combining the coding schemes 1 and 2 for internal and external memories as described in Memon et al. (2010). Scheme 1 included details to elicit true accounts, which include details derived from the senses, contextual details about spatial and temporal relationships and affective details or references to mood or emotional state. Internally made memories contained references to cognitive operations which is operationalised as anything insight the respondents head dealing with idiosyncratic memory and inferences about self or others (Memon et al., 2010). Scheme 2 used the same method to elicit truthful accounts, however cognitive operations were defined more broadly, including suppositions about sensory experience or reefing to descriptions of inferences made by the participants (Memon et al., 2010). These two schemes were combined to use the benefits of both coding schemes, with both schemes being close to the original RM framework, however, also looking at the quality of details. Internal memories consisted of cognitive operations which is anything inside the suspects head as e.g., reasoning, thought processes or cognitive processes. These indicate deception since truthful accounts should hold significantly less cognitive operations since truthful suspects can tell more external details as they report an experienced event. External memories consisted out of affective details, contextual details and external details. Affective details are anything that is related to emotion. Contextual details include details about location, spatial or

temporal details and relationships. External details are sensory information derived from all senses. See Table 1 for an overview of the coding.

Table 1

Coding for Internal and External Memories

Type of details	Example quote(s)
Internal memory	
Cognitive operations	"I thought it looked kind of suspicious", "I expected there to be a realtor"
External memory	
Affective details	"I did feel wrong to go to a house and look for people"
Contextual details	"a sign in the middle of the road"
External details	"I saw the door open" or "Nobody was there. It was completely empty"

Procedure

The participants signed an informed consent in an online survey. Participants were told about their right to withdraw and were provided with a cover story. The participant then got provided with a case providing a story to familiarize themselves with their condition (guilty vs. innocent, see Appendix C). The participants were then familiarised with the VR glasses and controllers. Guilty participants were instructed to rehearse stealing the training items by putting them into a white box, while innocent participants were not since they did not need to be able to manipulate items. Then the participants were instructed about their task again either being burglarising a house (i.e. guilty condition) or looking for a realtor inside an open house (i.e. innocent condition). Inside the house, guilty participants were told to steal anything they wanted without a time limit and to move out of the house when they thought that they had stolen enough and then the VR environment was turned off. Innocent participants were told to view the house for a potential house viewing. When the participants had completed the VR task, they returned to the computer to fill out questionnaires about experienced

cybersickness and presence in the VR environment which fell out of the scope of this study. When finished, the participants were asked to return to the experimenter.

The participants were taken to the interview room. During this walk, the experimenter disclosed to the participants that they were about to be interviewed by the police since a burglary had taken place in the house and a neighbour had seen a person looking like the participant in the neighbourhood during the time of the burglary. The participants were asked to convince the police of their innocence. In the interview room, the participants were asked to sit in front of the interviewer and the experimenter turned on the camera stating the number of the participant and exited the room. The interviewer performed the interview following the script (see material section or Appendix D). When the interview was finished the interviewer turned off the camera and asked the experimenter back into the room. The experimenter then took the participant back to the first room.

The participants were asked by the experimenter to fill in the last questionnaire of the study. This included the demographics, neurodiversity status self-reported rapport and self-reported cognitive load. When participants indicated to have no diagnosis during the demographics, they were asked to fill out the two described screeners for ASD and ADHD. When the participants had completed the questionnaire, they were asked to return to the experimenter. The experimenter debriefed the participant about the real reason for the study, making sure that the participants understood this and knew why the reason had been disclosed in the beginning. The participants were then thanked for their participation and asked to share the study with others.

Data Analysis

The data was analysed using R (version 4.4.2) using the following packages tidyverse, readxl, psych, dplyr, ggpubr, rstatix, performance, ggplot2, car, broom, emmeans and mediation. The data was tested for assumptions of normality and homogeneity. When assumptions would not be met, this is expected since it is an experimental study. ANOVAs can still be used since it is still a robust measure (Blanca et al., 2017). To test hypothesis 1 a two-way ANOVA was used with dependent variable external memories and independent variables neurotype and condition. To test sub-hypothesis 1a a mediation was used with dependent variables internal and external memories, as independent variables

neurotype and condition and as mediator cognitive load. The mediation was performed in two models. For the second sub-hypothesis a two-way ANOVA was used, with as dependent variables internal and external memories and independent variable guilty condition and neurotype.

Results

Descriptive Statistics

The dependent variables were correlated with each other, showing that internal and external memories correlated positively with each other, indicating that when more internal memories were mentioned, more external memories were mentioned and vice versa. Moreover, internal and external details were both positively skewed with most participants providing 5-6 external details and 1-2 internal details. In both groups a small number of participants named considerably more detials. Moreover, external memories was negatively correlated with cognitive load which indicates that when more external memories less cognitive load was experienced and vice versa. Cognitive load was approximately normally distributed, having a slight concentration at the midpoint. See Appendix H for an overview of histograms and density plots. On average, participants experienced moderate CL and named more external memories than internal memories (see Table 2).

Table 2

Mean, Standard Deviations and Correlations for Dependent Variables

	Mean	**SD**	**1.**	**2.**	**3.**
1. Cognitive Load	2.49	0.75	-		
2. External Memories	6.64	4.63	-0.29**	-	

3. Internal Memories	2.76	2.40	-0.09	.48***	-

Note. Reported correlations are corrected via Holm Corrections. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Main Analysis

External Memories

Before performing a factorial ANOVA, the assumptions were checked. Normality of residuals was tested using Shapiro-Wilk test finding that the residuals were not normally distributed, $W(97) = 0.905$, $p < .001$, see Appendix I for a Q-Q plot. Homogeneity of variance was assessed using Levene's test finding that the assumptions was met $F(3, 93) = 1.65$, $p = .184$. We still performed a factorial ANOVA to test for the main effects of condition (innocent vs. guilty) and autism (autism vs. neurotypical) on the number of external details mentioned in the investigative interview since ANOVAs are still a robust measure (Blanca et al., 2017). We observed statistically significant main effects for condition with a large effect size, $F(1,94) = 23.77$, $p < .001$, $\eta2 = 0.20$, 95% CI [0.09, 1.00]. This and all other interpretations of effect size are derived from Cohen (1988; 1992). This main effect indicated a higher number of external details for those in the innocent condition ($M = 8.67$, $SD = 4.93$) than in the guilty condition ($M = 4.56$, $SD = 3.20$). Although ASD individuals named slightly more external details ($M = 7.06$, $SD = 5.35$) than neurotypical individuals ($M = 6.19$, $SD = 3.73$), this effect was still not significant with autism explaining a small part of the variance in the outcome $F(1,94) = 1.16$, $p = .283$, $\eta2 = 0,04$, 95% CI [0.00, 1.00]. We observed a non-statistically significant interaction with a small effect size indicating that the effect of condition did not differ across groups $F(1,93) = 3.53$, $p = .063$, $\eta2 = 0,04$, 95% CI [0.00, 1.00]). Neurotypical innocent participants reported more external memories ($M = 7.42$, $SD = 4.01$) than neurotypical guilty participants ($M = 4.91$, $SD = 2.98$) which was the same for ASD innocent participants ($M = 9.88$, $SD = 5.49$) against ASD guilty participants ($M = 4.24$, $SD = 3.42$)

Internal Memories

Before performing a factorial ANOVA, again the assumptions were checked. Normality of residuals was tested using Shapiro-Wilk test finding that the residuals were not normally distributed, $W(97) = 0.919$, $p < .001$, see Appendix I for a Q-Q plot. Homogeneity of variance was assessed using Levene's test finding that the assumptions was met $F(3, 93) = 0.650$, $p = .585$. We performed a second factorial ANOVA to test for the main effects of condition (innocent vs. guilty) and autism (autism vs. neurotypical) on the number of internal details mentioned in the investigative interview. We observed statistically significant main effect for condition with a medium effect size, $F(1,93) = 48.02$, $p = .004$, $\eta 2 = 0.09$, 95% CI [0.02, 1.00]. This main effect indicated a lower number of internal details for those in the guilty condition ($M = 2.15$, $SD = 2.38$) than in the innocent condition ($M = 3.37$, $SD = 2.29$). We observed a statistically non-significant main effect for autism on internal details with a small effect size, $F(1,93) = 3.69$, $p = .410$, $\eta 2 = 0.03$, 95% CI = [0.00, 1.00], indicating no significant differences in number of internal details between ASD individuals ($M = 2.86$, $SD = 2.49$) and neurotypical individuals ($M = 2.66$, $SD = 2.32$). We observed a statistically non-significant interaction effect with a small effect size, indicating that the effect of condition did not differ across groups $F(1,93) = 13.99$, $p = 0.11$, $\eta 2 = 0,03$, 95% CI [0.00, 1.00]. Neurotypical innocent participants reported slightly more internal memories ($M = 2.88$, $SD = 1.92$) than neurotypical guilty participants ($M = 2.43$, $SD = 2.71$) which was more profoundly for ASD innocent participants ($M = 3.84$, $SD = 2.54$) against ASD guilty participants ($M = 1.88$, $SD = 2.05$)

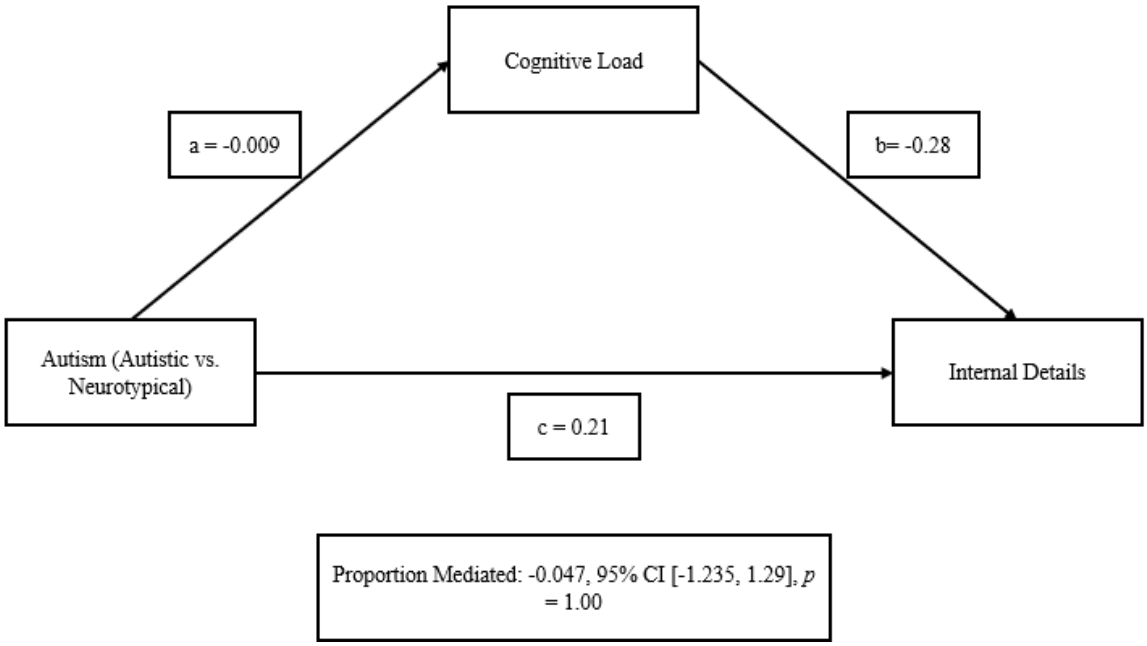
Mediation of Cognitive Load

Effect of Mediation on Internal Memories We performed a causal mediation using 5000 nonparametric bootstrap simulations to test if cognitive load mediates the relationship between neurotype (autism vs. neurotypical) and the number of overall details. The indirect effect (ACME) of cognitive load on internal details was not statistically significant, $\beta = -.009$, 95% CI [-0.133, 0.14], $p = .92$. Normally, the analysis would stop here, since this path is not significant. For the sake of completion, we decided to go further with the mediation analysis. The direct effect (ADE) of cognitive load on internal details was statistically not significant $\beta = 0.210$, 95% CI [-0.722, 1.15], $p = .66$. The total effect of autism on internal details was also non-significant $\beta = .20$, 95% CI [-0.746, 1.14], $p =$

.70. The proportion of the total effect mediated was also non-significant -0.047, 95% CI [-1.235, 1.29], p = 1.00. See Figure 2 for an overview of the mediation.

Figure 2

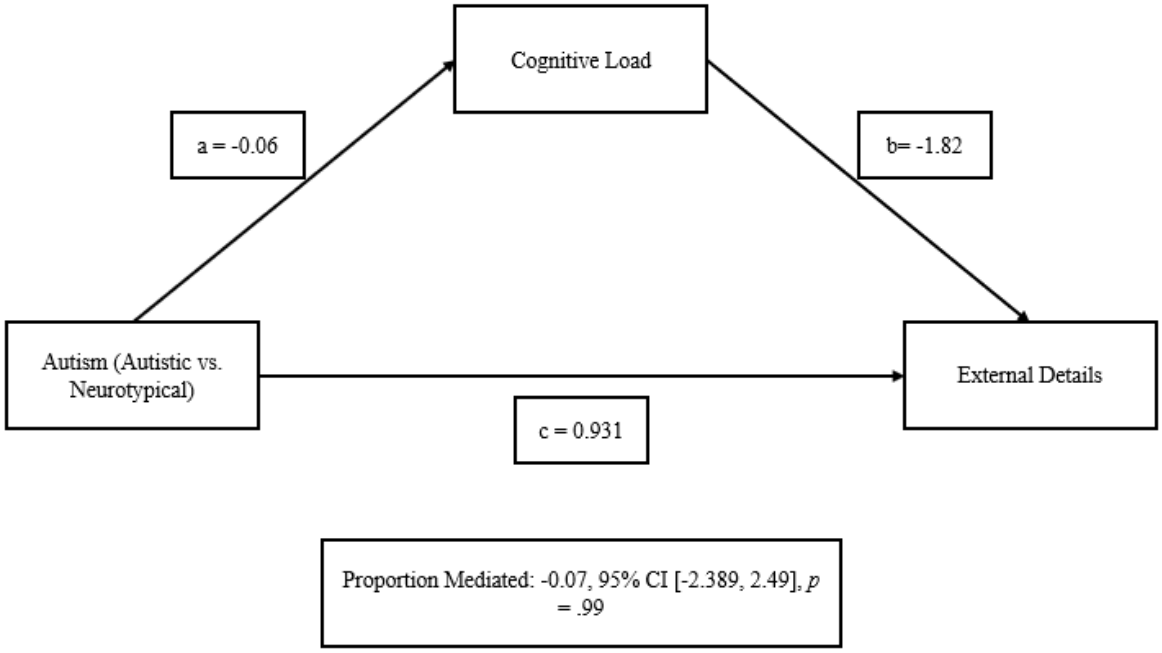
Mediation of Cognitive load on the Relationship between Neurotype and Internal Details



Effect of Mediation on External Memories Another causal mediation using 5000 nonparametric bootstrap simulations testing for the effect on external details showed that the indirect effect (ACME) of cognitive load on external details was not statistically significant β = -0.06, 95% CI [-0.657, 0.56], p = .82. Again, normally we would stop the testing here, since this path is non-significant. However, we decided to go further. The direct effect (ADE) of cognitive load on external details was statistically also not significant β = .93, 95% CI [-0.800, 2.74], p = .28. The total effect of autism on external details was also non-significant β = .87, 95% CI [-0.928, 2.72], p = .35. The proportion of total effect mediated was also non-significant -0.07, 95% CI [-2.389, 2.49], p = .99. See Figure 3 for an overview of the mediation.

Figure 3

Mediation of Cognitive load on the Relationship between Neurotype and External Details



Discussion

The main findings of this study are that, as expected, innocent participants provided significantly more external memories than guilty participants (Masip et al., 2005). However, against our expectations innocent participants did also mention more internal memories, which was thought to be higher in deceptive accounts (Masip et al., 2005). Next to this, against our expectations, we found no significant difference in the number of external and internal memories mentioned by ASD individuals or neurotypical individuals. Since ASD individuals have difficulties with the encoding of episodic memory (Norris & Maras, 2022), we expected that ASD individuals will provide fewer external memories which are based on episodic memory. Moreover, we expected that heightened cognitive load will mediate the overall mentioned memories in ASD individuals. However, we found no significant difference between the experienced cognitive load for ASD individuals and neurotypical individuals.

Neurotypical and ASD individuals both mentioned significantly more external memories in the innocent condition than in the guilty condition. This is as expected from the RM methodology which explains that truthful accounts have more external memories (Oberlader et al., 2016). However, we expected that this would be different for the accounts mentioned by ASD individuals, since ASD individuals have difficulties in encoding episodic memory and recall specific events (Norris & Maras, 2022). Therefore, we expected that ASD individuals will not mention more external memories and that

there will be a significant difference in the number of mentioned external memories between ASD individuals and neurotypical individuals, which was not the case in our study. However, there are some explanations for why this might not have been the case.

Since our sample consisted of higher educated young adults, we can assume that the individuals in the ASD group are cognitively and verbal able, which might explain why we see no differences. Diehl et al. (2006) analysed the difference in narratives between ASD children which are cognitively and verbal more able and neurotypical children, finding that there was no difference in length of narrative between ASD children and neurotypical children. In addition, in a longitudinal study, investigating narrative coherence over longer time, Almeida et al. (2024) found that cognitively and verbal able ASD individuals did not differ in effectively communicating their story even after longer time. However, ASD individuals showed a specific reduction in chronological coherence which was found to be related to directive prompts used (i.e., cued recall as “where/when did it happen” and focussing on event-related information using wh- questions). Nonetheless, Almeida et al. (2024) found that ASD individuals outperformed non-autistic individuals in richer descriptive markers providing richer details. This could explain why there is no difference in the number of details between the ASD and non-ASD group. Nonetheless, the increase in mentioned details is not very large, averaging one detail more mentioned by ASD individuals than neurotypical individuals.

On the other hand, against our expectations, cognitive load was not higher for ASD individuals than for neurotypical individuals in our study. Moreover, we did not find a mediating effect for cognitive load on the number of mentioned details. A possible explanation that cognitive load is not higher could be that the mock-interview had not the anticipated stressful effect on the participants. We did expect that due to the stress of the investigative interview situation, cognitive load would be increased. However, we also acknowledge that not only stress can increase cognitive load. Since the mock-interview used getting to know questions and was not provided in a stressful or pressuring way the participants might not have felt stressed. Interestingly, the rapport scores for both groups were found to be high without significant differences (Bouma, 2025). This indicates that ASD individuals and neurotypical individuals did not differ in their perceived rapport with the interviewer. Kang & Shen (2025) performed a study on the effect of using high-rapport building strategies by teachers on

the cognitive load of students finding that when rapport-building strategies were used this reduced the amount of experienced cognitive load of the students. Therefore, we can expect that rapport reduces cognitive load. Since the participants in our study overall reported a high rapport, this might indicate that the experienced cognitive load was lower.

Moreover, the interview was based on the SOM method, using open and funnelling questions, meaning that these might have included supporting cues for recall. For example, Alic et al. (2022) researched how to computationally identify funnelling and focusing questions on the classroom discourse, which does not tell us how funnelling questions lead to better recall, however, the results show that funnelling questions guide students towards more detailed questions. Moreover, research suggests that open-ended questions provide more detailed and accurate information than closed questions (e.g., Craig et al., 1999). Therefore, we can argue that the use of our interview strategy, did not only reduce cognitive load due to the experienced rapport, but also increased the amounts of details and improved recall by using a funnelling questions and open questions. This could indicate that we would only find a difference in the details in initial free recall and not when the interviewee is questioned using this approach.

Cognitive load was thought to disrupt the ability to recall episodic memory. However, cognitive load did not differ for ASD and neurotypical individuals. Therefore, the expected effect cognitive load might have on the number of mentioned details might be reduced or not be there is this situation. Moreover, since the cognitive control of individuals also might not have been disrupted as anticipated, providing new information or selecting new information should also not be more difficult. Therefore, the high rapport induced in combination with the used script without applying pressure to the interviewee might have unexpectedly reduced the cognitive load in a way that it did not influence the interviewee negatively. Therefore, this method of interviewing seems to be well-working for autistic individuals, however, we might consider that this also works well for neurotypical individuals. Therefore, this could mean that rapport does reduce cognitive load which shows the importance of building rapport with the interviewee.

Against our expectations, truthful accounts also mentioned more internal details. This is in line with the findings of Memon et al. (2010), which showed that there were more references to cognitive operations in truthful accounts. Other studies found varying results regarding the increase of internal memories in deceptive accounts (Memon et al., 2010). A reason for this could be that innocent suspect often are truthful, believing in the visibility of innocence when providing their account (Hartwig et al., 2007). This could result in the fact that innocent suspects also provide more internal details in their account. Moreover, according to Sporer (2004) idiosyncratic memory and cognitive processes support memory and processes of rehearsal and should therefore be considered as memory cues on self-experienced events rather than indicators of deception. This could explain why we found more internal memories in truthful accounts. However, also here the difference is rather small. Nonetheless, this also shows that internal memories might not be such a valid cue for detecting deception, as also mentioned by Masip et al. (2010).

Another explanation could be in the used screeners. The AQ-10 seems to be a reliable screener which is even proposed to use by the National Institute for Health and Care Excellence (NICE) (Hudson et al., 2024). Nonetheless, studies have provided criticism on the AQ-10, for example, that the AQ-10 has poor internal consistency in samples from the general population or has poorly fitting items (e.g., Jia et al., 2019). This research has led to questions about the psychometric rigor of the AQ-10 in the general population. Therefore, it could be that a part of the individuals which were assigned in the ASD group are wrongly screened positive for ASD. This could also explain that we did not find significant differences between ASD individuals and neurotypical individuals, since the difference between ASD individuals and neurotypical individuals is not as expected as neurotypical individuals were assigned to the ASD group. To test this an exploratory ANOVA was run, including only participants with an official diagnosis. However, the analysis did not find a significant other result than the mail analyses. A reason for this could be the reduced number of participants in the ASD group ($N = 18$) see Appendix J.

Limitations

This study used an experimental design which can reduce the ecological validity of this study since it is done in a lab. This reduces the generalizability of the results to the real-world context. Laboratory studies often do not incorporate social contexts (e.g., peer presence), social relationships or network which can shape or alter behaviour (Graham, 2017). Therefore, results may not be generalizable to settings outsight the lab, which might contribute to the replicability crisis (Agee et al, 2025). Nonetheless, the experimental design provides valuable insights allowing for a high control which makes the internal validity and reliability good. The study involved a type of interview like the SOM method, using two rooms likely similar to those in police interviews, trying to create a setting close to a real investigation environment. Nonetheless, when generalizing the results, we should be careful of the fact that we used a standardized script having no change of leading the interaction. Moreover, it is a mock-crime interview without the pressure to find a suspect or to show that your innocent which probably influences the behaviour of both the interviewer and the interviewee. Especially by using a VR environment and lab rooms for the interview, we could make sure that the experience and experimental set-up was the same for each participant.

For the interview, an interviewing script was used which reduced the ability to answer questions or build a more meaningful conversation with the interviewee from the interviewers' perspective. Therefore, a complete realistic interview scenario could not be simulated. However, by having an interview script it was made sure that there could be no differences between the way the interviewers conducted the interview. Having no interview script could have led to differences in the account of the interviewee induced by e.g., differing questions asked by the interviewer. Moreover, the interviewer might be able to probe answers better. Without these differences, we were able to accurately compare and gather the details mentioned in the interview which might have significantly differed when the interviewers would not have asked the same questions and used the same interviewing approach. This probably affected the findings because the provided details of truthtellers and deceivers might shave shown more or clearer differences, since deceivers would need to withhold information more and truthtellers would have more chances to provide more information (Hartwig et al., 2014; Hartwig et al., 2007).

The coding of the accounts was done by one researcher, which means that no interrater reliability could be assessed. This might reduce the reliability and quality of the coding, which could have impacted the results. However, for the coding the two coding schemes described in Memon et al. (2010) were used which provided reliable results researching RM. Therefore, we do not expect that this has impacted the results of this study significantly.

Further Research

From these limitations and the discussed results, further research is needed. We suggest that further research should be focused on the content of the accounts of the provided accounts. Accounts provided by cognitive and verbal able ASD individuals do not differ in length and richness of detail accounts or even exceed the accounts of non-ASD individuals. However, the accounts of ASD individuals might hold non-related details or lack chronological order. Therefore, the quality of details should be researched more deeply, to investigate the difference between the quality of details between accounts provided by ASD and non-ASD individuals. This could be done by automatically transcribing, assessing the relatedness of details and chronological order using an AI tool.

A more holistic approach could improve our understanding of how accounts are provided by ASD individuals. This research should focus on non-verbal and verbal communication which could provide valuable knowledge of what is indicative of deception and what is not. For example gaze aversion is often believed to be a sign of deception but often is done to regulate cognitive load (e.g., Doherty-Sneddon et al., 2012). Such (false) beliefs can affect how people judge others, as seen in Slijkhuis et al. (2025) showing that presuppositions of deceptive behaviours played a significant role in shaping indirect judgment accuracy. In an interview autistic individuals might show more gaze aversion trying to regulate cognitive load, however interviewers might wrongly believe that autistic individuals might be deceiving which is not the case. Results from such research can be used for training purposes for practitioners to improve how investigative interviews are conducted with vulnerable interviewees.

Research should be conducted on the cognitive load experienced in investigative on a broader sample consisting of not only academic participants. Moreover, in the research the stakes should be

increased to make the experience more realistic and increase the perceived stress and cognitive load. As mentioned, the sample consisted of cognitively and verbal skilled participants which might have reduced the cognitive load in both groups. Therefore, it would be interesting to research this effect with a more diverse sample. This could provide better understanding of the effect cognitive load might have on ASD individuals and the provided account. Moreover, combining a cognitive load measure with measures for stress could help to identify if ASD participants perceive more stress in the interview which could increase the cognitive load. For example. identifying the type or duration of stress could help to identify if this affects the experienced cognitive load or if the participants even feel stressed in such a situation.

Conclusion

To conclude, this paper aimed to answer this research question: *"Can Reality Monitory accurately detect deception in individuals with Autism Spectrum Disorders, considering the impairments in recall of episodic memory and cognitive load?".* We can say that our results showed that there is no significant difference in the number of provided details between ASD individuals and non-ASD individuals. Moreover, cognitive load seemed to not mediate the effect of ASD on the number of details. However, this does not directly mean that RM should be used to detect deception in ASD individuals as these results need to be interpreted with caution, since this could be due to the autistic sample not being typical. This research is one of the first studies researching the effectiveness of RM with ASD individuals and provides valuable insights in deception detection. However, more research is needed to provide more knowledge about the use of RM and accounts of ASD individuals to keep developing and improving the methods and strategies for detecting deception in the investigative interview.

Reference List

Abbe, A., & Brandon, S. E. (2013). The Role of Rapport in Investigative Interviewing: A Review. *Journal of Investigative Psychology and Offender Profiling*, *10*(3), 237-249. https://doi.org/10.1002/jip.1386

Adler, L., Kessler, R. C., Spencer, T., World Health Organization & Workgroup on Adult ADHD. (n.d.). *Adult ADHD Self-Report Scale (ASRS-V1.1) Symptom Checklist*. https://add.org/wp-content/uploads/2015/03/adhd-questionnaire-ASRS111.pdf

Agee, L. A., Fourtassi, A., & Monfils, M. (2025). Social context as a source of variability in the psychological sciences. *Frontiers in Human Neuroscience*, *18,* 1-7. https://doi.org/10.3389/fnhum.2024.1507010

Alic, S., Demszky, D., Mancenido, Z., Liu, J., Hill, H., & Jurafsky, D. (2022). Computationally Identifying Funneling and Focusing Questions in Classroom Discourse. *BEA 2022 - 17th Workshop on Innovative Use of NLP for Building Educational Applications, Proceedings*. https://doi.org/10.18653/v1/2022.bea-1.27

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward Brief "Red Flags" for Autism Screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 Cases and 3,000 Controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(2), 202-212. https://doi.org/10.1016/j.jaac.2011.11.003

Almeida, T.S., Yang, F., Zhang, H., & Michael, E. (2024) The Narrative Coherence of Autistic Children's Accounts of an Experienced Event in Response to Different Interviewer Prompts: A Longitudinal Study. *Journal of Autism and Development Disorders*. 1-16. https://doi.org/10.1007/s10803-024-06675-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

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

Bernieri, F. J. (1988). Coordinated movement and Rapport in teacher-student interactions. *Journal of Nonverbal Behavior*, *12*(2), 120–138. https://doi.org/10.1007/BF00986930

Blackhurst, T., Warmelink, L., Roestorf, A., & Hartley, C. (2024). The Brunswik Lens Model: a theoretical framework for advancing understanding of deceptive communication in autism. *Frontiers in Psychology*, *15*, 1-14. https://doi.org/https://doi.org/10.3389/fpsyg.2024.1388726

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29*(4), 552-557. https://doi.org/10.7334/psicothema2016.383

Bouma, D. (2025) *Nonverbal Behaviour as a Cue for Vulnerability: Examining ASD and ADHD Suspects in Investigative Interviewing* [Unpublished Master Thesis University of Twente].

Brewer, N., & Williams, K. D. (Eds.). (2005). *Psychology and law: An empirical perspective.* The Guilford Press.

Brewer, N., & Young, R. L. (2015) *Crime and Autism Spectrum Disorder.* Jessica Kingsley Publishers

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New York: Academic Press. https://www.utstat.toronto.edu/brunner/oldclass/378f16/readings/CohenPower.pdf

Cohen, J. (1992). A power primer. Psychological Bulletin, *112*(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Coutelle, R., Goltzene, M. A., Canton, M., Campiglia-Sabourin, M., Rabot, J., Bizet, É., Schoenberger, M., Berna, F., & Danion, J. M. (2021). Episodic Autobiographical Memory in Adults With Autism Spectrum Disorder: An Exploration With the Autobiographical Interview. *Frontiers in Psychiatry*, *11*, 1-10. https://doi.org/10.3389/fpsyt.2020.593855

Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. H. (1999). Interviewer Questions and Content Analysis of Children's Statements of Sexual Abuse. *Applied Developmental Science*, *3*(2), 77–85. https://doi.org/10.1207/S1532480XADS0302_2

Debbaudt, D. (2002) *Autism, Advocattes, and Law Enforcement Professionals Recognizing and Reducing Risk Situations for People with Autism Spectrum Disorders.* Jessica Kingsley Publishers.

Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *Journal of Abnormal Child Psychology*, *34*(1), 87-102. https://doi.org/10.1007/s10802-005-9003-x

Doherty-Sneddon, G., Riby, D. M., & Whittle, L. (2012). Gaze aversion as a cognitive load management strategy in autism spectrum disorder and Williams syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(4), 420-430. https://doi.org/10.1111/j.1469-7610.2011.02481.x

Duke, M.C., Wood, J.M., Bollin, B., LaBianca, J., & Scullin, M. (2018). Development of the Rapport scales for investigative interviews and interrogations (RS3i), Interviewee version. *Psychology, Public Policy, and Law, 24*(1), 64–79. https://doi.org/10.1037/law0000147

GOV.UK (June 21, 2023) *Achieving Best Evidence in Criminal Proceedings*. https://www.gov.uk/government/publications/achieving-best-evidence-in-criminal-proceedings

Graham, H. (2017). Why social contexts matter. *Addiction*, *112*(3), 396. https://doi.org/10.1111/ADD.13612

Gudjonsson, G. H. (2003) *The Psychology of Interrogations and Confessions: A Handbook*. John Wiley & Sons, Ltd.

Gudjonsson, G. H. (2006) The psychological vulnerabilities of witnesses and the risk of false accusations and false confessions. In A. Heaton-Armstrong E. Shepherd G. Gudjonsson D. Wolchover, *Witness testimony. Psychological, investigative and evidential perspectives* 61–75, Oxford University Press

Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic Use of Evidence During Investigative Interviews: The State of the Science. In Raskin, D. C., Honts, C. R., & Kircher, J. C. (Eds.)

Credibility Assessment: Scientific Research and Applications (pp. 3-36). https://doi.org/10.1016/B978-0-12-394433-7.00001-4

Hartwig, M., Granhag, P. A., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime and Law*, *13*(2), 213-227. https://doi.org/10.1080/10683160600750264

Hartwig, M., Granhag, P. A., Stromwall, L., Wolf, A. G., Vrij, A., & af Hjelmsäter, E. R. (2011). Detecting deception in suspects: Verbal cues as a function of interview strategy. *Psychology, Crime and Law*, *17*(7), 643–656. https://doi.org/10.1080/10683160903446982

Herrema, J.A. (2025) *The influences of Gradual and Late Disclosure of Evidence and Guilt and Innocence of a Suspect on Verbal Cues to Deception and Cognitive Load.* [Unpublished bachelor's thesis]. University of Twente

Herrington, V., & Roberts, K. (2012). Addressing Psychological Vulnerability in the Police Suspect Interview. *Policing*, *6*(2), 177-186. https://doi.org/10.1093/police/par057

Hoekendijk, J., & Beek, M.L.J. van (2015). The GIS-model: A Dutch approach to gather information in suspect interviews. *Investigative Interviewing: Research and Practice, 7*(1), 1-9.

Hudson, C. C., Bajwa, H. E., Beard, C., & Björgvinsson, T. (2024). Psychometric properties of the 10-item Autism Quotient in an acute psychiatric sample. *Research in Autism Spectrum Disorders*, *110*, 1-7. https://doi.org/10.1016/j.rasd.2023.102299

Jia, R., Steelman, Z. R., & Jia, H. H. (2019). Psychometric Assessments of Three Self-Report Autism Scales (AQ, RBQ-2A, and SQ) for General Adult Populations. *Journal of Autism and Developmental Disorders*, *49*(5), 1949-1965. https://doi.org/10.1007/s10803-019-03880-x

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67–85. https://doi.org/10.1037/0033-295X.88.1.67

Justus, S. A., Powell, P. S., & Duarte, A. (2021). Intact context memory performance in adults with autism spectrum disorder. *Scientific Reports*, *11*(1), 1-14. https://doi.org/10.1038/s41598-021-99898-2

Kang, H., & Chen, H. (2025) Effects of teachers' rapport-building strategies on EFL learners' cognitive load and computer-assisted language learning motivation. *Education and Information Technologies*, *30*, 3911-3952. https://doi.org/10.1007/s10639-024-12961-z

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245-256. https://doi.org/10.1017/S0033291704002892

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

Logos, K., Brewer, N., & Young, R. L. (2021). Countering Biased Judgments of Individuals Who Display Autism-Characteristic Behavior in Forensic Settings. *Human Communication Research*, *47*(3), 215-247. https://doi.org/10.1093/hcr/hqab002

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

Mackie, M. A., & Fan, J. (2016). Reduced Efficiency and Capacity of Cognitive Control in Autism Spectrum Disorder. *Autism Research*, *9*(3), 403–414. https://doi.org/10.1002/aur.1517

Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychology, Crime & Law*, *11*(1), 99-122. https://doi.org/10.1080/10683160410001726356

Memon, A., Fraser, J., Colwell, K., Odinot, G., & Mastroberardino, S. (2010). Distinguishing truthful from invented accounts using reality monitoring criteria. *Legal and Criminological Psychology*, *15*(2), 177–194. https://doi.org/10.1348/135532508X401382

National Institute of Mental Health (December, 2024) *Autism Spectrum Disorder*. https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd

Norris, J. E., & Maras, K. (2022). Supporting autistic adults' episodic memory recall in interviews: The role of executive functions, theory of mind, and language abilities. *Autism*, *26*(2), 513-524. https://doi.org/10.1177/13623613211030772

Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior*, *40*(4), 440-457. https://doi.org/10.1037/lhb0000193

Oleszkiewicz, S., & Watson, S. J. (2021). A meta-analytic review of the timing for disclosing evidence when interviewing suspects. *Applied Cognitive Psychology*, *35*(2), 342-359. https://doi.org/10.1002/acp.3767

Rispens, I., & van der Sleen, J. (2022). *Handleiding Verhoren van Kwetsbare Verdachten* [Manual Investigative Interviewing of Vulnerable Suspects]. Sdu BV.

Screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist in 1,000 Cases and 3,000 Controls. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(2), 202–212. https://doi.org/10.1016/j.jaac.2011.11.003

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental effort. *Annual Review of Neuroscience, 40*, 99-124. https://doi.org/10.1146/annurev-neuro-072116-031526

Slijkhuis, P. J. H., Watson, S. J., Stel, M., & Giebels, E. (2025, May 6). Gaze behaviour and deception detection: Insights from an eye tracking analysis of interviewers and observers. https://doi.org/10.31234/osf.io/dt7gh_v1

Sporer, S. (2004). Reality monitoring and detection of deception. In P. A. Granhag & L. A. Strömwall (Eds.), *Deception of detection in forensic contexts* (pp. 64–102). Cambridge: Cambridge University Press

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

U.S. Centers for Disease Control and Prevention (n.d.), *About Autism Spectrum Disorder*. https://www.cdc.gov/autism/about/index.html

van Sintemaartensdijk, I., van Gelder, J.-L., van Prooijen, J.-W., Nee, C., Otte, M., & van Lange, P. (2020). Mere presence of informal guardians deters burglars: a virtual reality study. *Journal of Experimental Criminology*, *17*(4), 657–676. https://doi.org/10.1007/s11292-020-09430-1

Vogels, J., Kramer, E., & Maes, A. (2014). How Cognitive Load Influences Speakers' Choice of Referring Expressions. *Cognitive Science*, *39*(6), 1396-1418. https://doi.org/10.1111/cogs.12205

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, *20*(1), 28-32. https://doi.org/10.1177/0963721410391245

Vrij, A., Hope, L., & Fisher, R. P. (2014). Eliciting Reliable Information in Investigative Interviews. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1). 129-136. https://doi.org/10.1177/2372732214548592

Vrij, A., Leal, S., & Fisher, R. P. (2018). Verbal deception and the model statement as a lie detection tool. *Frontiers in Psychiatry*, *9*, 1-6. https://doi.org/10.3389/fpsyt.2018.00492

Weiher, L., Watson, S. J., Taylor, P. J., & Luther, K. (2023). How multiple interviews and interview framing influence the development and maintenance of rapport. *Psychology, Crime and Law*, 1–25. https://doi.org/10.1080/1068316X.2023.2265527

Appendix

Appendix A

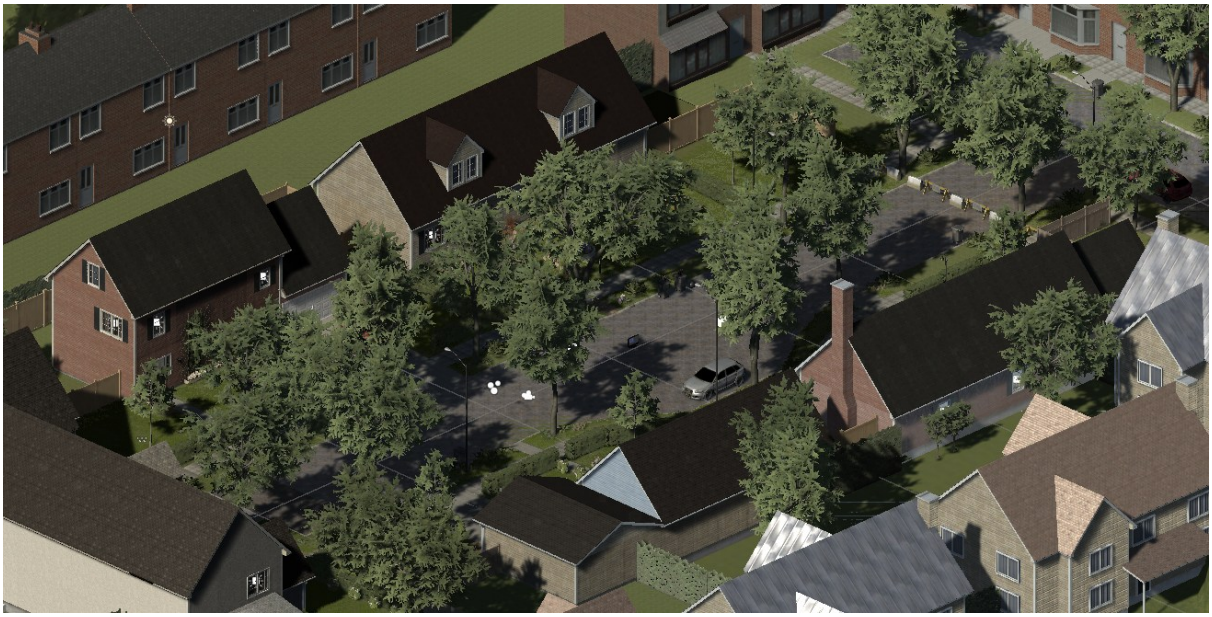
AI Statement

During the preparation of this work, I used ChatGPT to help brainstorming and for data analysis and error fixing in R. Moreover, I used the Formalizer Tool from Goblin Tools to improve the structure and clarity of the text. Next to this, I used Elicit and SCISPACE to improve my literature search. After using these tools/services, we thoroughly reviewed and edited the content as needed, taking full responsibility for the final outcome.

Appendix B

Overview of the VR Environment

Birdseye view of the neighbourhood



View of the open house

Practice items and white box on street

Appendix C

Case for Background Story Condition

Case – Burglar

Imagine yourself in the following situation: You are walking your usual route through a neighbourhood and have noticed multiple times that a pink house often has its front door open. You've scouted this house before, but today you decide to enter the house, as it appears no one is inside. Recently, you've had financial setbacks, making it difficult to buy groceries. Standing in the empty house, something shifts in your mind, and you decide to burglarize the house by taking valuable items.

Case – Visitor

Imagine yourself in the following situation: You're looking for a new home, and there is an open house event in a nearby neighbourhood. Since it's close by, you decide to walk there. You see pink house with an open front door and assume it must be the one for the viewing. As you're looking for a new home and the door is open for the open house, you decide to enter the house, thinking a realtor may be present.

Appendix D

Script for the Investigative Interview

Opening the interview

Hello, I would like to talk to you about a recent burglary in the neighbourhood. A neighbour reported seeing you in the area around [time of the VR study]. I would like to ask you a few questions to understand whether you know anything about the incidents.

I would like to go over some ground rules for today, alright? For your protection and for mine I will record this, so we get a full account of what was said today. Also, when you're talking, I'm not going to interrupt you, and I hope you can do the same for me. If you have any questions, please ask me. I might ask you to repeat some things because I want to make sure I understand everything. Does that make sense?

The main purpose here is to get as much information as possible. So, it is important that you tell me everything in as much detail as possible without leaving things out. This is important because I wasn't there, so I don't know what happened. Do you understand?

Personal oriented interview

Before we start with specific questions, I would like to get to know you better:

1. Where do you live? With whom do you live?
2. Do you have a job? How long have you worked there, and what are your main tasks?
3. What do you study? Do you enjoy it, why?
4. What does a typical week look like for you?
5. Do you have any hobbies? Could you describe for me what you like about your hobbies?

Now I will ask you some more specific questions. You may have already answered them, but if that is the case, please answer them again.

As mentioned, you are here to talk about a recent burglary in the neighbourhood. A neighbour reported seeing you in the area around [time of the VR study] and we have some question about that.

Case specific interview

1. Please tell me in as much detail as possible, why you were in the neighbourhood?
2. Describe in as much detail as possible what you were doing while you were in the neighbourhood?
3. What did you see or notice while you were in the neighbourhood? Please tell me, in as much detail as possible.
4. What other details can you tell me that might be important? Is there anything else you noticed that you haven't mentioned?
5. Did you see anyone else in the neighbourhood? If so, please describe them and the interactions you had.
6. A neighbour noticed a person looking like you entering a house. Can you please explain why you entered the house?

a. ENTERED THE HOUSE: Please tell me in as much detail your purpose of entering the house?
b. DID NOT ENTER THE HOUSE: If you did not enter the house, how do you explain a neighbour seeing a person looking like you entering the house?

i. ALIBI: Okay, I understand you didn't enter the house, can you tell me what you did that day?

7. Is there anything else you would like me to know?

[After questioning] Thank you very much for your time and information. I have everything we need for now. You can wait here, and the research assistant will be in to see you.

Appendix E

Screeners for Neurodiversity

Screener for ASD

	Please tick one option per question only:	Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
1	I often notice small sounds when others do not				
2	I usually concentrate more on the whole picture, rather than the small details				
3	I find it easy to do more than one thing at once				
4	If there is an interruption, I can switch back to what I was doing very quickly				
5	I find it easy to 'read between the lines' when someone is talking to me				
6	I know how to tell if someone listening to me is getting bored				
7	When I'm reading a story I find it difficult to work out the characters' intentions				
8	I like to collect information about categories of things (e.g. types of cars, types of bird, types of train, types of plant etc)				
9	I find it easy to work out what someone is thinking or feeling just by looking at their face				
10	I find it difficult to work out people's intentions				

Scoring: Only 1 point can be scored for each question. *Score 1 point for Definitely or Slightly Agree on each of items 1, 7, 8, and 10. Score 1 point for Definitely or Slightly*

Disagree on each of items 2, 3, 4, 5, 6, and 9. If the individual scores **six or above**, consider referring them for a specialistic diagnostic assessment.

Screener for ADHD

		Never	Rarely	Sometimes	Often	Always
1	How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done?					
2	How often do you have difficulty getting things in order when you have to do a task that requires organization?					
3	How often do you have problems remembering appointments or obligations?					
4	When you have a task that requires a lot of thought, how often do you avoid or delay getting started?					
5	How often do you fidget or squirm with your hands or feet when you have to sit down for a long time?					
6	How often do you feel overly active and compelled to do things, like you were driven by a motor?					

Scoring: Only 1 point can be scored for each question. *Score 1 point for Sometimes, Often or Always on each of items 1, 2 and 3. Score 1 point for Often or Always on each of items 4, 5 and 6.* If the individual scores **four or above**, consider referring them for a specialistic diagnostic assessment.

Appendix F

Self-Report Questionnaire Cognitive Load

Cognitive Load Scale for Processes and Verbalizing the Memory

	1 (Strongly Disagree)	2 (Disagree)	3 (Neither Agree nor Disagree)	4(Agree)	5(Strongly Agree)
I found it difficult to explain the order of events while being interviewed.					
I noticed that I shared more information than I wanted to when answering the police officer's questions.					
I found it difficult to answer the police officer's questions as fully as I wanted to.					
I felt that it took me a long time to think through how to answer the police officer's questions.					
Sometimes after giving an answer, I wished I could go back and restart or change my answer.					
I found it difficult to translate the mental story I created into an actual statement.					

Cognitive Load Scale for Retrieving the Memory

	1 (Strongly Disagree)	2 (Disagree)	3 (Neither Agree nor Disagree)	4(Agree)	5(Strongly Agree)
When answering the police officer's questions, I found it difficult to remember the details I wanted to talk about.					
I found it difficult to remember my overall story when answering the police officer's questions.					

I seemed to forget what I already told the police officer and what I did not.

It required all my concentration to answer the police officer's questions.

It was hard work to remember what I wanted to say.

Scoring: The total cognitive load score was calculated by summing the points from all the statements. *Strongly Disagree* is assigned 1 point and *Strongly Agree* is 5 points. A maximum of 55 points could be achieved. For analyses, the mean scores of each participant were used. A higher mean score indicated a higher level of cognitive load.

Appendix G

Self-Report Questionnaires Rapport

RS3i (Duke et al., 2018)

Below you will find various statements on a scale from 1 = "Strongly disagree" to 5 = "Strongly agree". Please pick the answer that best describes the way you feel.

Rapport Scale for Investigative Interviews and Interrogations (RS3i) Interviewee Version

	1 (Strongly Disagree)	2 (Disagree)	3 (Neither Agree nor Disagree)	4(Agree)	5(Strongly Agree)
I think the police officer was generally honest with me.					
The police officer was skilful during the interaction.					
The police officer seemed to respect my knowledge.					
The police officer and me have our culture in common.					
The police officer performed expertly during the interaction.					
I think that the police officer can generally be trusted to keep their word.					
The police officer and me probably share the same ethnicity.					
The police officer really listened to what I had to say.					
I was motivated to perform well during the interaction.					
I feel I can trust the police officer to keep his word to me.					
The police officer made an effort to do a good job.					
The police officer acted like a professionals.					
The police officer paid careful attention to my opinion.					

The police officer and me got along well during the interaction.

The police officer and me worked well together as a team.

The police officer probably shares my culture.

I wanted to do a good job during the interaction.

The police officer was attentive to me.

Communication went smoothly between the police officer and me.

The police officer was interested in my point of view.

I felt committed to accomplishing the goals of the interaction.

Scoring: The total self-reported rapport score was calculated by summing the points from all the statements. *Strongly Disagree* is assigned 1 point and *Strongly Agree* is 5 points. A maximum of 105 points could be achieved. For analyses, the mean scores of each participant were used. A higher mean score indicated a higher level of self-reported rapport.

Interaction Scale (Bernieri, 1988)

Please rate the interaction between yourself and the police officer on each of the characteristics listed below on a scale from 1 = "Strongly disagree" to 5 = "Strongly agree". Please pick the answer that best describes the way you feel.

Rapport Scale for Interaction Between the Interactants

The interaction seemed…

	1 (strongly disagree) (1)	2 (disagree) (2)	3 (neither agree nor disagree) (3)	4 (agree) (4)	5 (strongly agree) (5)
...well-coordinated. (1)					
...cooperative. (2)					
...harmonious. (3)					
...positive. (4)					
...friendly. (5)					

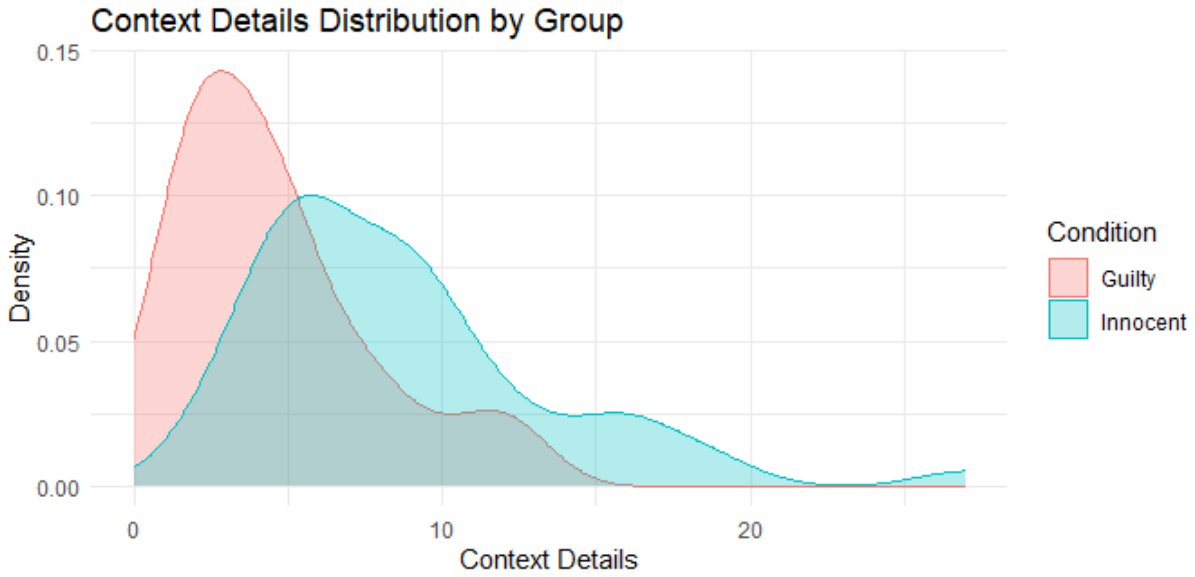
...respectful. (6)

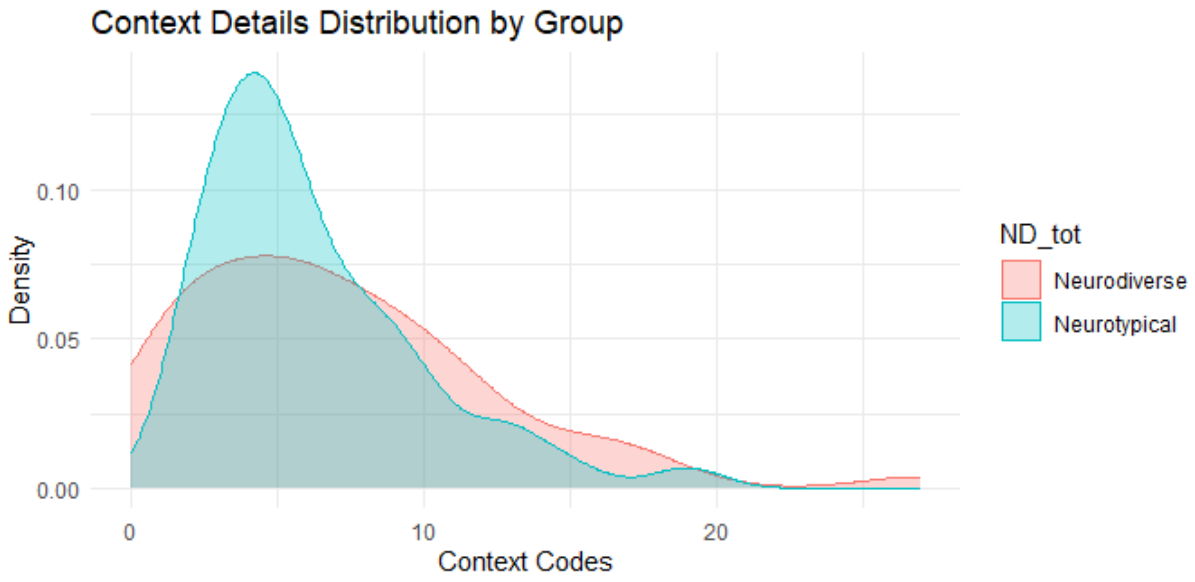
…attentive. (7)

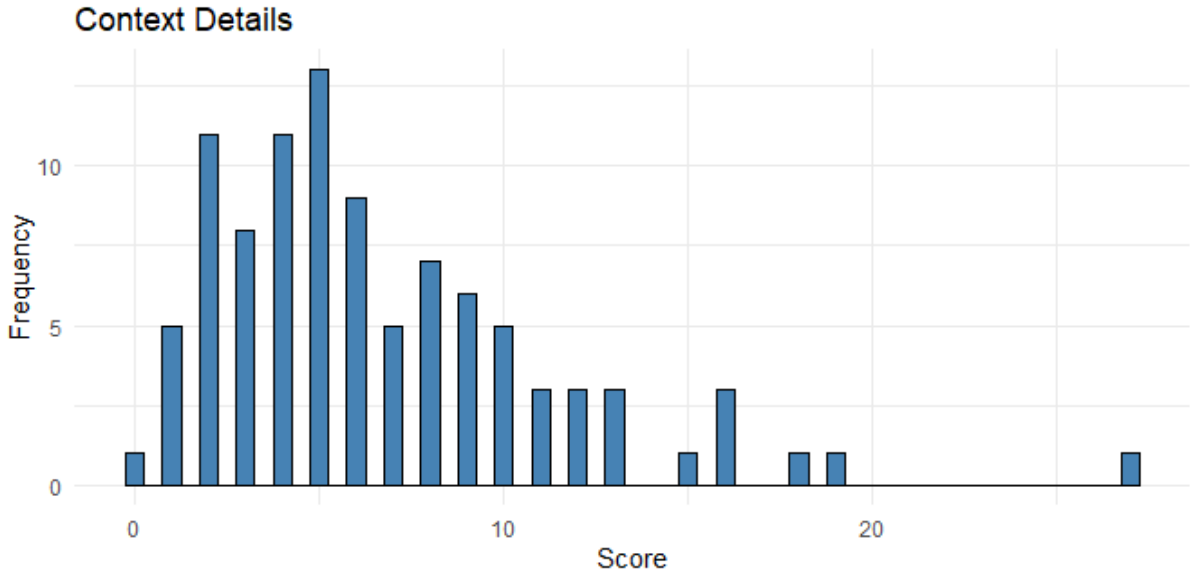
Scoring: The total self-reported rapport score was calculated by summing the points from all the statements. *Strongly Disagree* is assigned 1 point and *Strongly Agree* is 5 points. A maximum of 35 points could be achieved. For analyses, the mean scores of each participant were used. A higher mean score indicated a higher level of self-reported rapport.

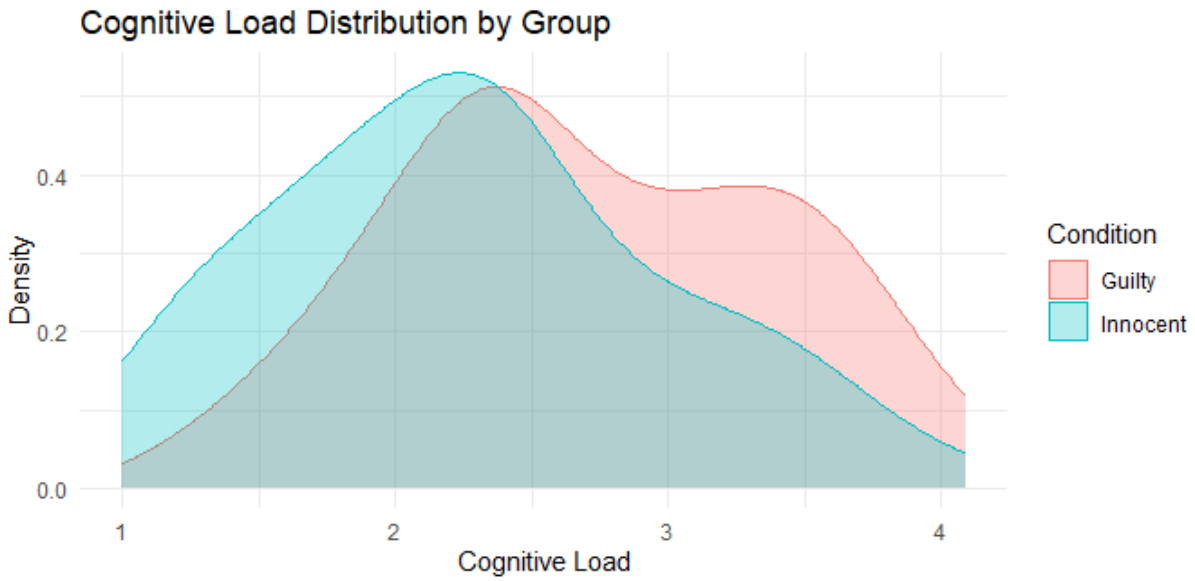
Appendix H

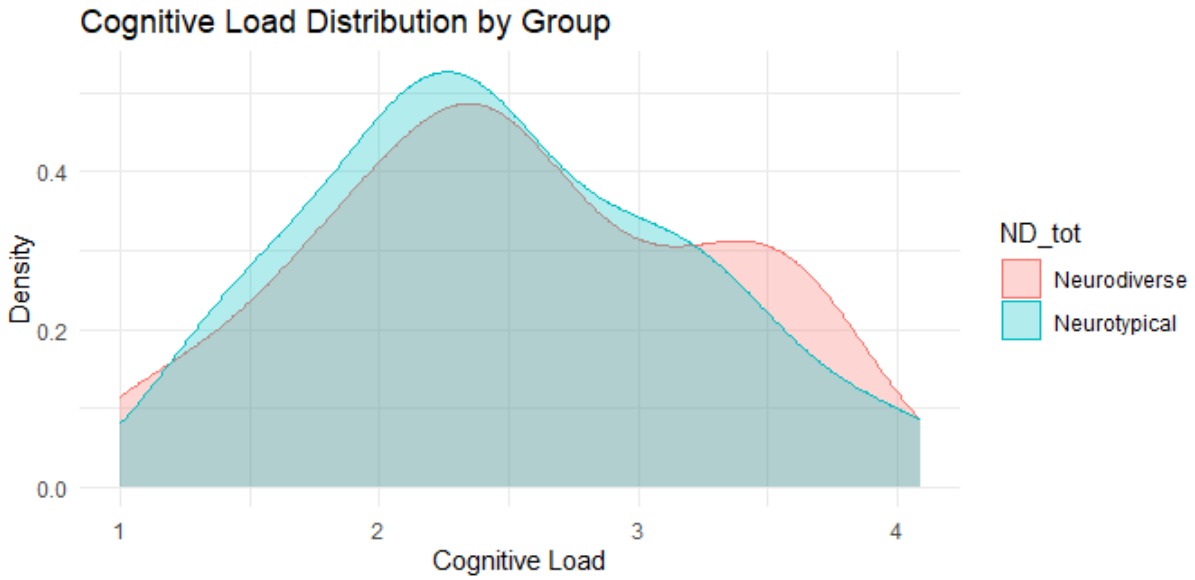
Histograms and Density Plots of Dependent Variables

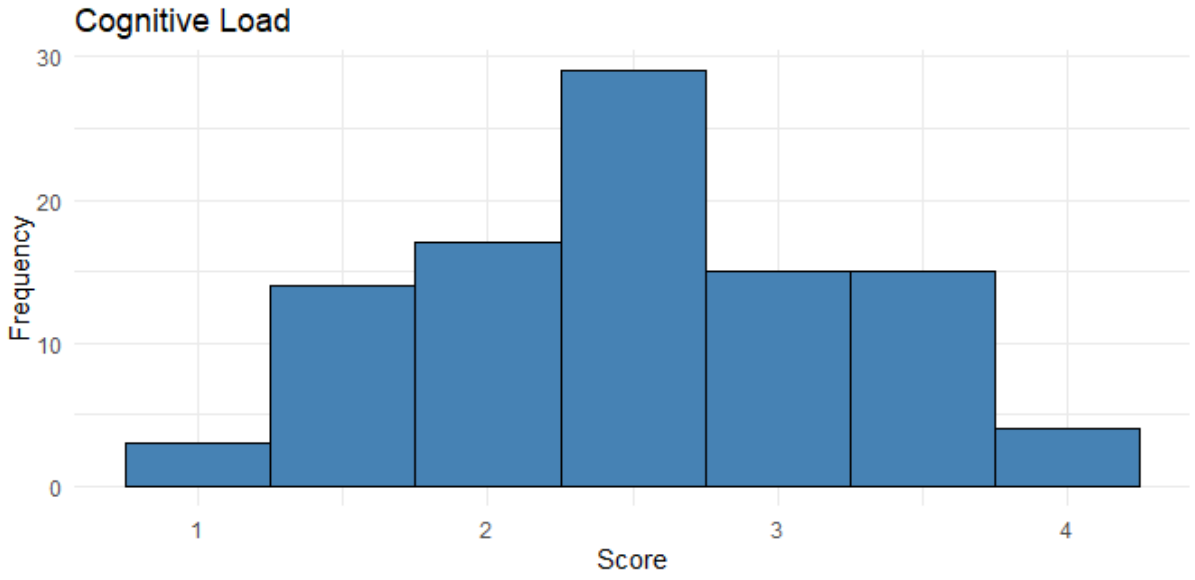


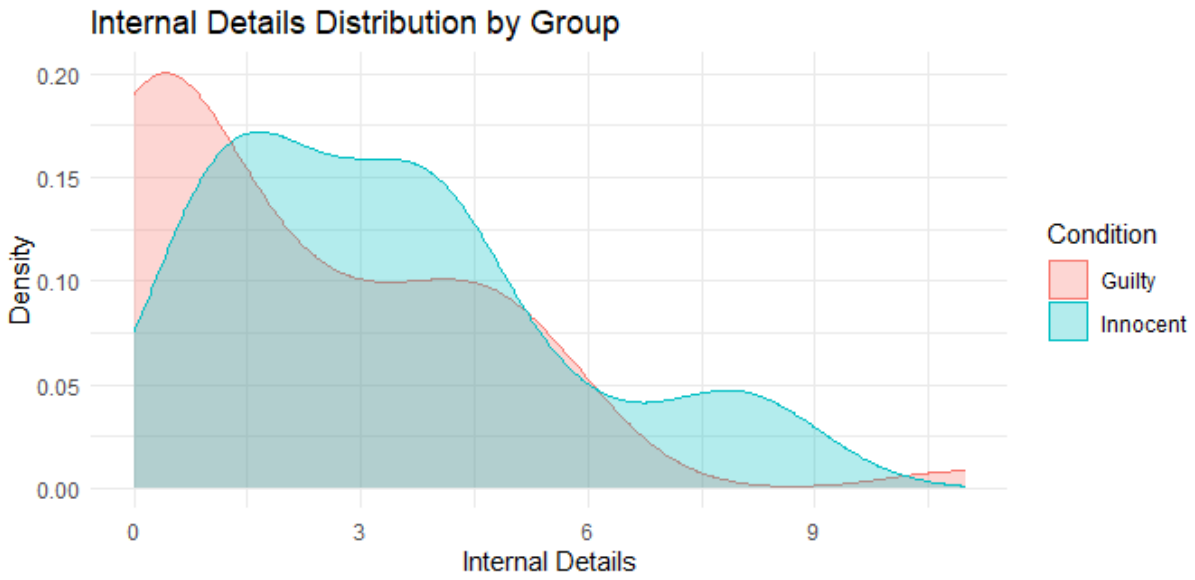


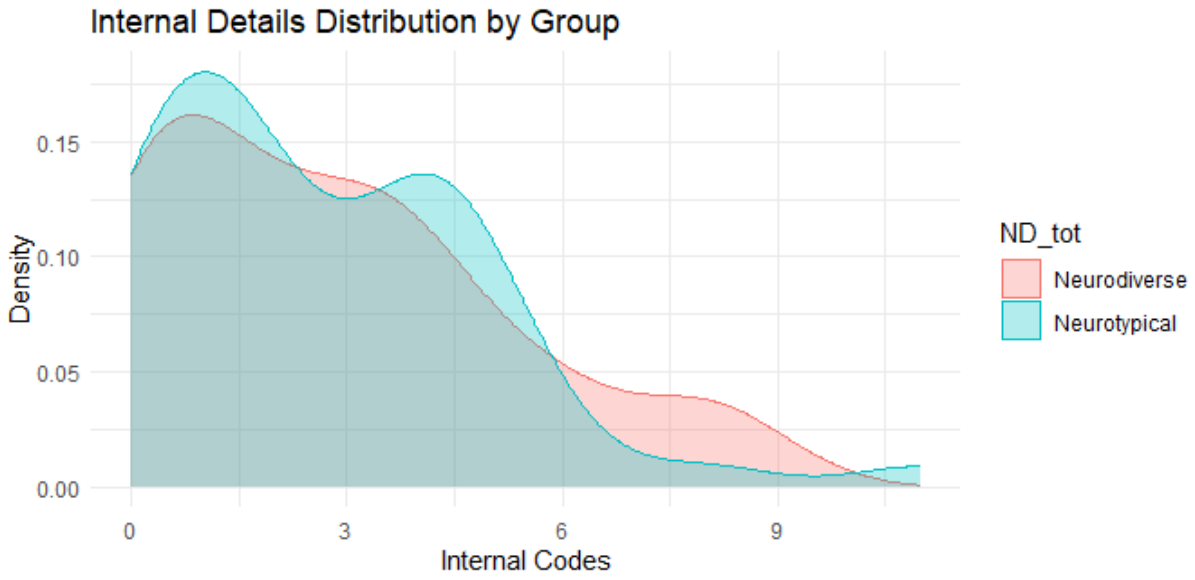


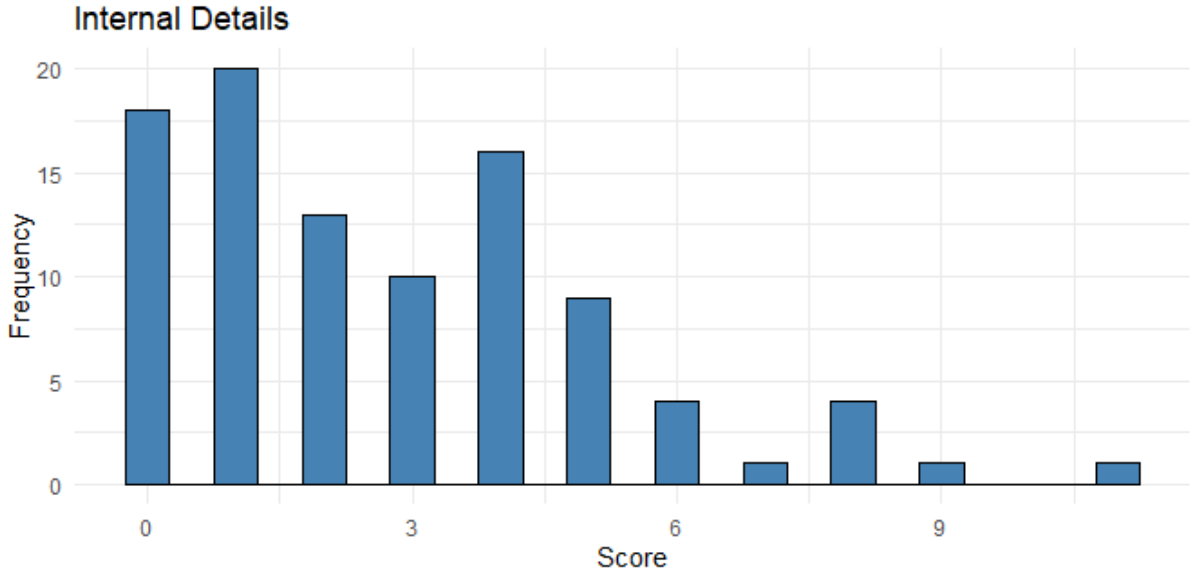

Cognitive Load Distribution by Group
Density
0.0
0.2
0.4
1
2
3
4
Cognitive Load
Condition
Guilty
Innocent


Cognitive Load Distribution by Group
Density
0.0
0.2
0.4
1
2
3
4
Cognitive Load
ND_tot
Neurodiverse
Neurotypical


Cognitive Load
Frequency
0
10
20
30
1
2
3
4
Score

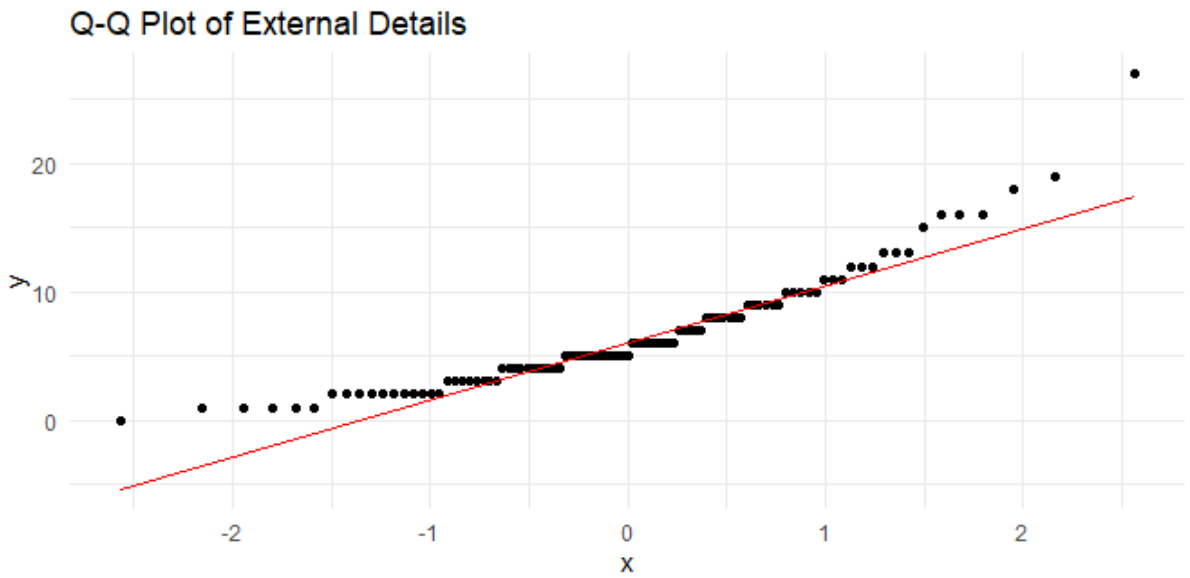

Internal Details Distribution by Group
Density
0.20
0.15
0.10
0.05
0.00
0
3
6
9
Internal Details
Condition
Guilty
Innocent

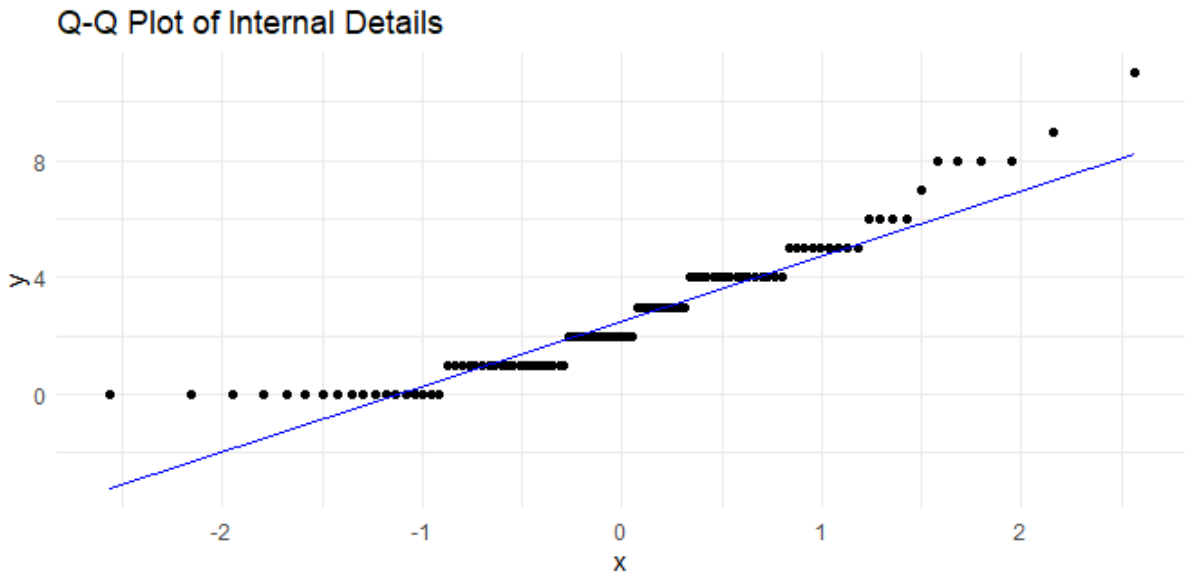

Internal Details Distribution by Group
Density
0.15
0.10
0.05
0.00
0
3
6
9
Internal Codes
ND_tot
Neurodiverse
Neurotypical


Internal Details
Frequency
20
15
10
5
0
0
3
6
9
Score

Appendix I

Q-Q Plots for External and internal Memories





Appendix J

Exploratory ANOVAs

External Memories

Before performing the ANOVAs we made a new dataset in which we only included participants with an official diagnosis in the ASD group. We then performed a factorial ANOVA to test for the main effects of condition (innocent vs. guilty) and autism (autism vs. neurotypical) on the number of external details mentioned in the investigative interview. We observed statistically significant main effects for condition with a small effect size, $F(1,53) = 48.01$, $p < .04$, $\eta 2 = 0.08$, 95% CI [0.00, 1.00]. This main effect indicated a higher number of external details for those in the innocent condition ($M = 7.7$, $SD = 3.54$) than in the guilty condition ($M = 4.70$, $SD = 2.91$). ASD individuals named slightly less external details ($M = 6.17$, $SD = 2.89$) than neurotypical individuals ($M = 6.31$, $SD = 3.75$), this effect was still not significant with autism explaining a small part of the variance in the outcome $F(1,53) = 5.01$, $p = .501$, $\eta 2 = 0{,}009$, 95% CI [0.00, 1.00]. We observed a non-statistically significant interaction with a small effect size indicating that the effect of condition did not differ across groups $F(1,53) = 4.00$, $p = .548$, $\eta 2 = 0{,}007$, 95% CI [0.00, 1.00]). Neurotypical innocent participants reported more external memories ($M = 7.65$, $SD = 3.93$) than neurotypical guilty participants ($M = 4.91$, $SD = 3.05$) which was the same for ASD innocent participants ($M = 7.86$, $SD = 2.04$) against ASD guilty participants ($M = 3.80$, $SD = 2.17$)

Internal Memories

We performed a second factorial ANOVA to test for the main effects of condition (innocent vs. guilty) and autism (autism vs. neurotypical) on the number of internal details mentioned in the investigative interview. We observed statistically non-significant main effect for condition with a small effect size, $F(1,53) = 4.00$, $p = .400$, $\eta 2 = 0.01$, 95% CI [0.00, 1.00]. This main effect sill indicated a lower number of internal details for those in the guilty condition ($M = 2.44$, $SD = 2.65$) than in the innocent condition ($M = 3.07$, $SD = 1.98$). We observed a statistically non-significant main effect for autism on internal details with a small effect size, $F(1,53) = 0.012$, $p = .963$, $\eta 2 = 0.00004$, 95% CI = [0.00, 1.00], indicating no significant differences in number of internal details between ASD

individuals (*M* = 3.08, *SD* = 2.19) and neurotypical individuals (*M* = 2.69, *SD* = 2.37). We observed a statistically non-significant interaction effect with a small effect size, indicating that the effect of condition did not differ across groups F(1,53) = 1.18 *p* = 0.64, η2 = 0,004, 95% CI [0.00, 1.00]. Neurotypical innocent participants reported slightly more internal memories (*M* = 2.91, *SD* = 1.95) than neurotypical guilty participants (*M* = 2.45, *SD* = 2.77) which was more profoundly for ASD innocent participants (*M* = 3.57, *SD* = 2.15) against ASD guilty participants (*M* = 2.40, *SD* = 2.30)