Detecting Suspicious Behaviour: The Influences of Autism and Theory of Mind

By

Zoe Michael
BPsych (Hons)

Thesis

Submitted to Flinders University
for the degree of

Doctor of Philosophy (Clinical Psychology)

College of Education, Psychology, and Social Work

25th January 2024

TABLE OF CONTENTS

Summary

Despite a lack of substantial empirical support for the claim that Theory of Mind (ToM) difficulties pose a unique risk factor for criminal vulnerability in autistic defendants, it is frequently cited as a mitigating factor in defence arguments and expert testimony. It has been argued that ToM difficulties may lead people to become embroiled in criminal activity due to an inability to recognise cues to suspicious behaviour in others (Brewer & Young, 2015). Although previous research has found evidence to suggest that ToM may be related to the detection of suspicious behaviour (Brewer et al., 2018; Brewer et al., 2023), there is currently no evidence to suggest that, on a group level, autistic individuals are less likely to detect cues to suspicious behaviour within interactions (Brewer et al., 2023). This study sought to further investigate the proposition that that (1) ToM difficulties are associated with trouble detecting and responding to suspicious behaviour within interactions, and that (2) both ToM difficulties and trouble detecting suspicious behaviour are more likely to present in autistic individuals than non-autistic individuals.

To test these hypotheses, I conducted three studies. Study 1 involved the development and extensive evaluation of a novel paradigm called the Suspicious Activity Paradigm (SAP) using large samples of non-autistic adults. The SAP was created to provide a more ecologically valid measure of the detection of suspicious behaviour than tasks used in previous research (Brewer et al., 2018; Brewer et al., 2023). Study 1 demonstrated that (1) the paradigm presented a viable way of capturing an array of responses detailing how people respond to different problematic situations, (2) responses could be reliably coded within a series of categories using a comprehensive protocol, and (3) the patterns of responding across participants were relatively stable which indicated the sensitivity of the paradigm to the proposed cues to suspicion.

Study 2 then extended upon Study 1 by introducing comparison measures of ToM, autistic traits, and verbal ability for a preliminary indication of the SAP's relationship to each measure. This study generally found no relationships between ToM or autistic traits and the SAP but highlighted significant restrictions in variability on the ToM measure. Furthermore, it indicated that verbal ability had a meaningful association with SAP performance.

The third and final study involved adding an autistic sample to investigate autistic-non-autistic group differences and further explore associations between the SAP and ToM, autistic traits, and verbal ability. It also included additional measures of gullibility, social vulnerability, and interpersonal trust as potential markers of concurrent and divergent validity. Study 3 demonstrated no group-level difference in the ability to detect and respond to suspicious behaviour. It also showed that many autistic participants performed at or near ceiling on the ToM measure, which constrained the ability to detect meaningful relationships between ToM and the SAP. Nevertheless, comparison of very high and very low ToM scores demonstrated that those with poor ToM were less likely to respond adaptively or report suspicion than those with very high ToM, irrespective of diagnosis. Study 3 did not provide independent confirmation of the SAP's convergent validity using the proposed validity markers and highlighted a need for ongoing validation in future research.

Taken altogether, my findings suggest that ToM is an important social-cognitive feature that should be considered when discussing vulnerability to criminal involvement, regardless of diagnosis. In addition, although it should not be assumed to impact all autistic individuals equally, due to the increased likelihood of prominent ToM difficulties in autistic adults, ToM should be thoroughly investigated when autistic adults appear in court as a defendant. Further research is suggested to explore the influence of intellectual ability and other characteristics of autism, and how these interact with ToM to increase vulnerability.

Declaration

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

The studies for this thesis were supported by an Australian Research Council Discovery grant to N. Brewer & R. L. Young.

Additionally, I confirm that I received an Australian Government Research Training Program Scholarship to support the completion of this thesis.

Signed

Date 27/10/23

Acknowledgments

Neil, it has been a pleasure to learn from you over the past six years. Your guidance, wealth of knowledge, and wit have been invaluable throughout this experience. I have grown immensely as a researcher under your watch and I'm forever grateful that you forewent a relaxing retirement to continue your important work and take me on as a student.

Robyn, thank you for inviting me into your lab and for your moral support throughout the process. The weekly lab meetings have always been a valuable time for connection, reflection, and a laugh.

To Mum, Dad, and Jackson, I wouldn't be here if you hadn't shown me that it was possible. There's nothing like a bit of healthy competition to inspire achievement! Your love, support, and understanding of the PhD slog carried me through. I am so fortunate to have had you all to lean on.

To Michael, thank you for your unwavering support and patience over the last few years. I am so, so grateful for all you've done for me and excited for our next chapter in the big smoke.

Finally, to my friends - particularly Cam and Lara - I can't thank you for helping me to get *any* work done, but I can thank you for all the laughter, joy, and lake walks. RIP PK.

CHAPTER 1

Introduction

The ability to take the perspective of others, often referred to as Theory of Mind (ToM), is fundamental to successful social interaction. Observing and understanding the world from various points of view allows us to communicate effectively, forge and maintain relationships, and manage conflict. Conversely, ToM difficulties present barriers to understanding, interpreting, and inferring the thoughts, emotions, intentions, and motivations of others. Within our complex social world, these barriers can significantly impact wellbeing in a variety of contexts, including relationships and employment. Additionally, ToM difficulties may have disastrous consequences if an individual fails to recognise malicious intent within an interaction. At present, very little is known about the impact of ToM difficulties on potential vulnerability to criminal involvement arising from manipulation or coercion.

Autism spectrum disorder (ASD) is a complex neurodevelopmental condition that affects approximately 1 in 59 individuals (Baio et al., 2018). While heterogenous in its presentation, the condition is characterised by two key criteria: (1) marked deficits in verbal and non-verbal social communication across a range of contexts, and (2) restricted and repetitive patterns of behaviour or interests (American Psychiatric Association, 2013). Autistic traits typically become evident in the early developmental period but will rarely manifest in the same manner, or to the same degree, from one individual to the next. Since its conception as a disorder (Kanner, 1943), an association between ASD and deviant or criminal behaviour has endured in both the media and psychological literature, despite the absence of conclusive evidence. The present research investigates whether certain characteristics associated with autism, namely ToM difficulties—that is, difficulties in taking the

perspective of others—are associated with difficulty detecting suspicious behaviour, and how that might impact vulnerability to involvement in criminal activity.

Prevalence of Criminality in Autism

A perceived association between autism and violent crime has likely been fuelled by sensationalist media reporting (Jones & Harwood, 2009; Sellers, 2018; Young et al., 2012) and early case-reports that led researchers to speculate that many violent offenders in custody may have Asperger's Syndrome (Mawson et al., 1985). Empirical research regarding the prevalence of ASD in offending populations has revealed mixed results. Approaches to the collection of prevalence data have varied in the literature, with three general measures predominating: (1) the prevalence of offending among people with autism, (2) the prevalence of autism within offender populations, and (3) rates of offending behaviour among those with ASD in the community. King and Murphy's (2014) comprehensive review of autism within the criminal justice system focussed on autistic individuals who demonstrated offending behaviour. Twenty-two studies were analysed according to approaches (1) and (2), as well as psychiatric co-morbidity in people with autism who offend, types of offence committed, and vulnerabilities of autistic people within the criminal justice system. Estimates for the prevalence of autism among offender populations ranged from 3% to 27%, while offending behaviour in people with autism ranged from 2.74% to 26% (or 48% when including self-reported criminal activity). They concluded that the outcomes reported in the 22 studies involved were difficult to compare due to widely differing methodologies and research foci. Small samples ($N < 40$ in 17 of the 22 studies) in mostly biased populations such as forensic psychiatric facilities, sub-par diagnostic protocols, and varied measures of offending highlighted a need for further rigorous research. These conclusions were later echoed in a review by Rutten et al. (2017). A further limitation with the prevalence research lies in the

frequency of offenders who receive their first diagnosis upon entering the criminal justice system, making generalisations to the wider community difficult (Mouridsen, 2012).

Recently, Collins et al. (2023) conducted an updated systematic review to address (a) whether King and Murphy's (2014) recommendations regarding methodological improvements that might enhance the rigour of the relevant literature had been fulfilled since their original review, and (b) whether knowledge and understanding of autism within the criminal justice system had improved. Like King and Murphy (2014), they concluded that the quality of the literature remained poor because of methodological limitations including "poor generalisability, unrepresentative samples, lack of matched comparison samples, reliance on retrospective data collection and lack of standardised instruments" (p. 3169). They also reported a minor over-representation of autism in the criminal justice system indicated by higher reported prevalence rates of autism in the criminal justice system (between 0.2% and 62.8%) than in the general population. However, estimates between studies varied considerably and the reliability of these data was questionable given the methodological limitations. They conceded that the findings of the review were limited by the poor quality of the included articles, and again highlighted a need for more rigorous research with a particular focus on improving training and awareness of autism within the criminal justice system.

Taken altogether, the picture remains unclear. Several studies suggest that autism may be disproportionately overrepresented in offenders, relative to its prevalence in the population, particularly for crimes such as arson (Allely, 2019; Cashin & Newman, 2009; Mouridsen, 2012) and sexual offences (Kumagami, 2006). Recently, however, Yu et al. (2021) analysed epidemiological data from the United States over a 15-year period to track charges, outcomes, and recidivism rates in young adults with autism. Autistic individuals were compared with a non-autistic community control and a sample of intellectually disabled

individuals and results revealed that, overall, groups did not differ in charges, outcomes, or rates of recidivism, and autistic young adults were not overrepresented in either the juvenile or adult justice systems. The view that there is no clear link between autism and criminality has been shared by a number of researchers (e.g., Brewer & Young, 2015; Browning & Caulfield, 2011; Mouridsen, 2012; Railey et al., 2020).

Theory of Mind and Autism

Regardless of whether autistic individuals are over-represented in terms of commission of crime, numerous case studies support the possibility that there may be characteristics of autism that render individuals vulnerable to involvement in crime (Baron-Cohen, 1988; Brewer & Young, 2015; Howlin, 2004). For example, Brewer and Young (2015) detailed the arrest of 'Kosta', a young autistic man who pointed a water pistol toward a fast-food employee at the request of some acquaintances he believed to be 'mates'. Kosta did not believe that his actions constituted a robbery, as "he did not think the cashier would believe the gun to be real and could not understand why the cashier was so distressed" (p. 126). Kosta's case demonstrated a failure to understand and appreciate the impact of his actions on others and an apparent inability to anticipate the intentions of the men who had baited him.

One theoretical explanation for criminal vulnerability in autistic individuals such as Kosta is that autistic individuals' difficulties with social communication reflect an impaired ability to take the perspective of others. This is demonstrated in a lesser ability to understand, interpret and infer the thoughts, emotions, intentions, and motivations of others from their verbal and non-verbal behaviour. Such difficulties are thought to indicate ToM difficulties, a concept which has dominated discussions about the social-cognitive characteristics of autism for several decades (Baron-Cohen et al., 1985). Successful social interaction relies upon an implicit understanding of the thoughts, emotions and motivations of one another, and acknowledgement that they differ between people. Thus, it is believed that an impairment in

the ability to understand and infer the mental states of others presents a barrier to effective communication in dynamic social interactions.

Measuring ToM in Adults

A vast literature regarding ToM difficulties has emerged over the years, particularly in the context of child development (Baron-Cohen et al., 1985; Doherty, 2008; Wellman et al., 2001). The traditional, gold-standard measure of ToM, the 'false-belief' task, requires respondents to demonstrate an understanding that individuals' beliefs about the world can diverge from reality. First-order false belief tasks require respondents to infer beliefs about another person, while second-order false belief tasks require respondents to reason about what one person thinks of another person's thoughts. It is well documented that autistic children perform more poorly on false belief tasks than non-autistic children (Baron-Cohen, 2001; Baron-Cohen et al., 1985; Wellman et al., 2001), with the latter expected to demonstrate proficiency on first-and second-order tasks by the ages of four and six, respectively (Doherty, 2008). Yet, examinations of ToM across the lifespan reveal that false-belief tasks have failed to produce the consistent group differences between autistic and non-autistic individuals that are seen in children. It has been argued, however, that the lack of discrimination between autistic and non-autistic individuals in adults provides insufficient evidence to support the idea of an intact ToM (Baron-Cohen et al., 1997) and, consequently, measures of 'higher-order' ToM were devised to better reflect the complexities of the construct and reduce ceiling effects. For example, while autistic adults have demonstrated the ability to recognise differential mental states at the false-belief level, they may still exhibit clear difficulties in interpreting more subtle mental state cues such as facial expressions or non-literal speech. Furthermore, while false-belief tasks are representative of the initial dichotomous 'pass or fail' conceptualisation of the ToM construct, higher-order ToM tasks

are typically mastered at varying ages and ability can be expressed on a continuum as complexity increases.

To combat the discrepancy between proficiency on basic ToM tasks and real-life difficulties, researchers have attempted to refine the measures used to detect ToM. Methodological approaches to measuring ToM in adults have differed, with some targeting implicit ToM, or the ability to take the perspective of others without conscious appraisal, and others targeting explicit ToM, which requires a deliberate evaluation of others' mental states. Implicit ToM tasks often include eye-tracking measures embedded within false-belief tasks to measure anticipatory looking, and these paradigms have shown that implicit ToM can exist independently of task instruction (Schneider et al., 2014). For instance, Senju (2012) found that adults who were able to effortlessly pass explicit false-belief tasks were unable to spontaneously anticipate actions by demonstrating appropriate gaze patterns.

Many other tasks are considered to measure explicit ToM and require a respondent to actively evaluate the perspective of a character or entity. One such task, The Strange Stories Task (Happé, 1994), requires participants to respond to a series of short vignettes of everyday conversational situations that involve non-literal speech such as white lies, sarcasm, and double bluff. There are also seven 'physical' situations that do not involve any mentalising ability, with these expected to reveal similar performance across groups. For each vignette, accompanying questions test individuals' comprehension of the situation. Similar measures depicting realistic scenarios in video format have since been introduced: for example, the Adult-Theory of Mind (A-ToM; Brewer et al., 2017) and The Awkward Moments Test (Heavey et al., 2000).

Other tasks present a hybrid of implicit and explicit methodology. The Reading the Mind in the Eyes Test or 'The Eyes Test' (Baron-Cohen et al., 2001; Baron-Cohen et al., 1997) requires participants to infer basic mental states such as happy, sad, and afraid, as well

as ‘complex’ mental states such as reflective, scheming, arrogant, and planning, from a series of photographs of the eye region. Similarly, the Frith-Happé animations (Abell et al., 2000) requires participants to respond to three sets of geometric animations that represent random, goal-directed, and ToM related motions. The random and goal-directed animations represent either purposeless or physical movements, respectively, while the ToM motions depict interactions in which one triangle reacts to another’s mental state. Participants are asked to describe what they think has happened in each animation.

The variety of presentations also extends to the requirements of responding, with some tasks allowing time to deliberate and consider a response (e.g., Strange Stories), some providing forced-choice options (The Eyes Test, Frith-Happé animations, The Awkward Moments Test), and others requiring naturalistic, time-restricted responding (A-ToM). Measures with a lack of time constraint have been criticised for allowing participants to use analytical reasoning to “hack out” solutions, rather than provide an indication of the person’s ability in real-life contexts (Frith, 1994). The inclusion of naturalistic stimuli and open-ended response formats has become a point of emphasis since research has suggested that the disparity between success on laboratory-based tasks and real-life difficulties may indicate that impairments are most likely to present when manufactured situations mimic the demands of real life (Frith, 2004; Frith et al., 2003; Ponnet et al., 2008; Roeyers et al., 2001).

Evidence for ToM Difficulties in Autistic Individuals

While the data regarding ToM in children reveal a relatively clear distinction between autistic and non-autistic individuals’ abilities, research using adult populations has produced more varied results. Many studies using the aforementioned tasks have demonstrated clear deficits in autistic mentalising ability when compared with non-autistic participants (Kleinman et al., 2001), and equivalent ability on the ‘physical’ elements of tasks that do not require mentalising (Baron-Cohen et al., 1997; Brewer et al., 2017). Other studies, however,

have demonstrated group differences on both the ToM and non-social facets of a task (Heavey et al., 2000; Zalla et al., 2009). Problematically, when multiple ToM measures are compared within studies, the results are sometimes contradictory. For instance, Spek et al. (2010) found that autistic participants performed more poorly than their non-autistic counterparts on the Strange Stories and Faux-pas (Stone et al., 1998) tasks, but did not observe any group differences on The Eyes Test. Similarly, Brewer et al. (2017) were able to demonstrate group differences across the A-ToM, Strange Stories, and Frith-Happé animations, although the Strange Stories also differentiated groups based on the non-social items. Furthermore, the Frith-Happé animations revealed only a weak significant difference between autistic and non-autistic participants which became nonsignificant after controlling for verbal intelligence (Brewer et al., 2017).

Additionally, the literature to date is not without methodological limitations. Many adult ToM measures have not been thoroughly evaluated or standardised for use as formal assessment tools as this was not their intended purpose. Many findings are based on very small sample sizes, and the use of different measures in research has been inconsistent, so the ability to review and compare results is limited. For instance, while the Strange Stories task has been widely used, modifications have often been made to the content (White et al., 2009), number of scenarios (Jolliffe & Baron-Cohen, 1999) or method of presentation (O'Hare et al., 2009), rendering a cohesive review of the relevant literature difficult. Furthermore, while ToM is considered to operate independently of intelligence, verbal intelligence has been identified as a significant contributor to success on many ToM tasks (Happé, 1995). Despite this, many studies do not control for verbal intelligence (Demurie et al., 2011; Gray et al., 2011). Conversely, the need to control for verbal intelligence means that the literature is often restricted to a subset of autistic individuals that do not have an intellectual disability and, as intellectual disability frequently co-occurs in individuals with autism (Matson & Shoemaker,

2009), findings are not necessarily representative of the population as a whole. Regardless, the literature to date has indicated an apparent inability of at least some autistic adults to perform at a similar level to their non-austistic peers on a range of ToM measures, echoing clinical and anecdotal evidence of significant social-communication difficulties throughout the autistic population.

Taken together, the available evidence suggests that, on average, autistic individuals perform more poorly than their IQ-matched non-autistic counterparts on measures of ToM (Baron-Cohen, 2001; Brewer et al., 2017; Kimhi, 2014; Peñuelas-Calvo et al., 2019; Yirmiya et al., 1998). However, whether these results are indicative of a core deficit in ToM remains questionable. Core-deficit theories propose that all profiles within a diagnostic category can be explained by a ubiquitous impairment (Astle & Fletcher-Watson, 2020). Gernsbacher and Yergeau (2019) argue that it is inappropriate for researchers to have accepted the assertion that autistic people lack a ToM as the theory relies on refutable evidence. They contend that the literature fails to demonstrate adequate specificity, universality, replication, and validity. That is, ToM deficits are not unique to autistic individuals, not all autistic individuals demonstrate deficits in ToM, many results supporting the ToM deficit theory have failed to replicate, and there is limited support for the predictive and convergent validity of ToM measures (but, for exceptions, see Brewer et al., 2017, 2019; Young & Brewer, 2020). Indeed, given the variability that can be seen within groups across various ToM measures, the suggestion of a ubiquitous deficit in autistic individuals appears dubious. Similarly, it could be argued that the inconsistent group differences do not support the notion of a universal impairment. Nevertheless, it is undisputed that the spontaneous nature of dynamic social interactions presents a significant barrier for many autistic individuals and the adult ToM tasks require participants to make inferences about other people's intentions, actions, and emotions, providing an indication of their ability in these areas.

ToM and Criminal Vulnerability

Difficulties in taking the perspective of others can greatly impact autistic individuals' ability to navigate important facets of everyday life, such as gaining employment and maintaining relationships. In the workforce, excellent interpersonal skills and adherence to social norms are common conditions of employment, and research has consistently indicated that autistic individuals are subject to poor employment outcomes despite a strong desire to work and contribute to their community (Gal et al., 2015; Hendricks, 2010; Taylor et al., 2015). ToM difficulties may cause problems for individuals with knowing how to present 'well' in resumés and interviews and, once employed, understanding instructions. In a similar vein, ToM difficulties can create significant barriers to gaining and maintaining interpersonal relationships. When examining romantic relationship status and relationship satisfaction among autistic individuals, Strunz et al. (2017) found that most participants who were single at the time of assessment endorsed an interest in relationships, but commonly reported barriers such as "contact with others being too exhausting", "fear of not fulfilling partner expectations", and "not knowing how to behave in a relationship" as reasons for remaining single. Such concerns are similarly applicable to the maintenance of friendships and peer relationships, which have also been reported as relatively low in prevalence (Mazurek, 2014; Orsmond et al., 2004).

While the negative impact of ToM difficulties on quality of life is clear, less is known about whether such difficulties render autistic individuals more vulnerable to criminal involvement. It must be noted that poor ToM cannot be solely blamed for the commission of any crime, as several personal and environmental factors must converge to create a particularly unfortunate circumstance. Still, it would appear from the outcomes observed in several case studies that autistic individuals may find themselves in problematic interactions and fail to detect subtle verbal and non-verbal cues that indicate the malicious intent of

others. Therefore, the present research will investigate whether ToM difficulties may render autistic individuals vulnerable to naïve criminal involvement, given certain situational conditions.

It is important to reiterate that the available prevalence data suggest most autistic individuals are law-abiding citizens. In fact, cognitive rigidity is a common feature of autism and may serve as a protective factor against criminal vulnerability in individuals that have a strong moral code or insistence on sameness (Howlin, 2004; Mouridsen, 2012). For instance, autistic individuals who are 'rule-bound' may be especially unwilling to humour others if it requires flexibility from their usual routine or contradicts their view of how the world should be. In these cases, it is possible that an individual would be less likely to fall victim to manipulation than a non-autistic individual with more advanced adaptive skills. Nevertheless, it appears that certain characteristics of the autistic individual may increase their criminal vulnerability in certain situations.

Currently, the literature regarding the relationship between ToM and criminal vulnerability is in its infancy. However, some studies have explored the possibility that criminal offenders may have ToM difficulties. Recently, Karoğlu et al. (2021) conducted a systematic review examining ToM ability in offenders and non-offenders. They reported generally mixed results regarding group differences in ToM but emphasised limitations in the quality of the studies included. For example, several studies did not control for important confounding variables such as cognitive ability or contained small sample sizes. Furthermore, an array of ToM measures with varying levels of psychometric quality were used. Nevertheless, when examining only the highest quality studies, they found that the research suggested offenders may have impairments in higher-order ToM compared to non-offenders.

Some recent studies have also approached the issue by examining the relationship between ToM and the ability to detect suspicious or dodgy behaviour. To my knowledge,

four studies currently exist in this area. Brewer et al. (2018) reported the first study of ToM and the ability to detect suspicious behaviour. In their paradigm, participants listened to several audio vignettes of unfolding scenarios, some of which culminated in a crime, while others did not. Participants were instructed to indicate when they believed something suspicious was occurring or, conversely, whether nothing suspicious was occurring, by pressing buttons (i.e., "seems a bit dodgy", "certain that something is very dodgy", "no longer seems dodgy" and "nothing was dodgy throughout") on a piece of electronic equipment. After controlling for IQ, a weak negative relationship was found between the ToM components of the Frith-Happé animations (Abell et al., 2000) and time taken to detect suspicious activity. Importantly, as the study was devised as a preliminary exploration of the relationship between ToM and the ability to detect suspicious activity, it only involved non-autistic adults, thereby limiting the variance in ToM ability. Furthermore, the suspicion recognition latency data were based on very few trials, with participants only listening to 5 crime scenarios. Given the characteristic variability of latency data due to factors such as attentional fluctuations and individual differences in speed-accuracy operating characteristics, estimates based on so few trials are likely to be unstable, again constraining the ability to detect any meaningful statistical relationships. Therefore, although the results from Brewer et al. (2018) were a promising start, the nature and strength of the relationship remains largely unknown.

In another study, Williams et al. (2018) investigated the relationship between ToM and the ability to detect deception. Across two experiments, participants watched a series of videos which depicted people being accused of cheating in a research task. Participants were then required to make a categorical judgement about whether the person was a liar or truth-teller. In the first experiment, which used only non-autistic college students, they found that ToM and deception detection were unrelated, yet deception detection was related to a

measure of autistic traits. In the second study, when comparing autistic and non-autistic samples, they found that autistic individuals performed significantly worse on the deception detection task. Unfortunately, a measure of ToM was not included in this second component of their study and, therefore, it is unclear whether ToM difficulties underpinned the poorer deception detection performance. It is possible that the range of ToM ability in the original non-autistic sample was too narrow to detect a relationship between ToM and deception detection. Alternatively, the ToM measures used (Frith-Happé animations and Reading The Mind in the Eyes task) may not have been sufficiently sensitive.

Most recently, Brewer et al. (2023) employed a signal detection theory approach to investigate the relationship between ToM and the detection of suspicious behaviour. They included 136 written vignettes from 34 short stories that either culminated in a crime or did not. The vignettes were first piloted to obtain a rating of how suspicious or 'dodgy' each vignette was according to a dichotomous yes/no rating and corresponding confidence judgement (6-point scale from "sure not dodgy" to "sure dodgy"). Once a range of signal strengths to dodginess was confirmed, the vignettes were presented in random order to a second sample across two testing sessions. Participants were required to provide both a categorical response (i.e., yes or no) to indicate whether anything dodgy was going on, and then a rating of how sure they were about that judgement, presented on a 6-point scale ranging from sure it is dodgy to sure it is not dodgy. Contrary to the current hypotheses, they found no meaningful group differences between autistic and non-autistic individuals on the d' index of discrimination of dodgy cues, despite significantly poorer ToM in the autistic sample. They also found no group difference in response bias, or the tendency to report dodginess, indicated by the c index. When examining the confidence ratings using receiver operating characteristic (ROC) analyses, no group differences emerged, nor did a significant relationship between autistic traits and dodginess detection. However, when the groups were

combined, ToM and verbal ability were independently associated with the statistical indices of dodginess discrimination (i.e., *d'* etc.). These results suggest that ToM difficulties at an individual level, regardless of an autism diagnosis or presence of autistic traits, are associated with an impaired ability to detect suspicious behaviour. Brewer et al. (2023) concluded that case studies depicting criminal vulnerability in autistic defendants should not be dismissed but, rather, that there are likely additional considerations such as IQ, contextual factors, prior life experience, and other individual differences that may make important contributions to level of vulnerability. They also pointed out the inherent limitation in ecological validity of their particular signal detection paradigm, which differs significantly from real social interactions, and the need for further examination with more complex and dynamic stimuli. Thus, although these studies presented important preliminary findings, the limitations and paucity of research in this area necessitates further investigation.

A fourth related study by Young and Brewer (2020) investigated the relationship between ToM and the ability to provide information that would extricate oneself from police investigation when erroneously accused of a crime. Participants completed the A-ToM and Frith-Happé Animation tasks before listening to several scenarios in which they were instructed to assume the position of a person accused of committing a crime. After each scenario, participants were asked to provide any information which would serve to eliminate them as a suspect. Each scenario contained four possible details of extricating information that would serve as an effective alibi. The study found that autistic participants performed more poorly than non-autistic participants on both measures of ToM and the extrication task, with ToM mediating the relationship between group and extrication ability. While it did not directly address the identification of suspicious activity, the findings of this study inspired the inclusion of questions regarding behaviour in the present research, to further investigate the relationship between ToM and the ability to extricate oneself from problematic interactions.

Objectives

The research documented in this thesis sought to build upon the findings by Brewer et al. (2018), Williams et al. (2018), Young and Brewer (2020), and Brewer et al. (2023) by further investigating potential relationships between autism, ToM, and the ability to detect suspicious behaviour. Specifically, it examined (a) whether autistic adults would have greater difficulty detecting and responding to suspicious behaviour within interactions than non-autistic adults, and (b) whether ToM, irrespective of diagnosis, would be related to the ability to detect and respond to suspicious behaviour. That is, would people with autism and/or poor ToM find it difficult to identify that something suspicious was occurring, and would they also be able to respond to the situation in a way that would extricate them from possible criminal activity?

Considering the finding from Brewer et al. (2023) that ToM and verbal ability independently predicted dodginess discrimination, while autism diagnosis and autistic traits did not, the current project investigated whether that finding would replicate with a more ecologically valid measure for the detection of suspicious behaviour. To address this question, it was necessary to create a measure suitable for the purposes of this research. Thus, a novel paradigm for measuring the detection of suspicious behaviour was developed and evaluated across a series of three large-sample studies.

CHAPTER 2

Development and Piloting of the Suspicious Activity Paradigm

The purpose of the first component of the research program was to develop and pilot a new paradigm for the detection of suspicious activity. The paradigm used by (Brewer et al., 2018) measured latency to detect cues to suspicious behaviour and the signal detection approach by Brewer et al. (2023) used written vignettes. The present paradigm was designed to provide a more ecologically valid measurement of the ability to detect and respond to suspicious activity. This approach sought to incorporate and extend the evidence from Young and Brewer (2020) that some autistic individuals find it difficult to extricate themselves from problematic situations. Moreover, this approach permitted investigation of whether an individual's behaviour may be incongruent with their appraisal of the situation: for example, the possibility that an individual may recognise that a situation is suspicious but acquiesce to a request regardless.

Development of the Paradigm

To investigate the ability to detect and respond to suspicious activity, a novel paradigm—labelled the Suspicious Activity Paradigm (SAP)—was developed and subjected to extensive piloting involving two different phases. The SAP is a computerised task consisting of several scenarios that either depict an unfolding crime or a neutral storyline. Each crime scenario involves a character of interest being invited by another person, either a friend or stranger, to engage in criminal activity. Importantly, they were naïve to the criminal intentions of the other character. Scenarios were based on written vignettes used by Brewer et al., (2018) and were adapted to use current tense with verbal dialogue between characters, and third-person narration. The scenarios were then recorded as an ensemble where possible using iPhone 8 Voice Memos, or independently by each voice actor, and edited together using Adobe Creative Cloud Audition.

In each scenario, participants are required to take the perspective of one character. At several intervals throughout the scenario participants are probed as to how they would behave or react at that point if they were in that character's situation and, importantly, why they would react in that way. Scores on the task reflect both an ability to detect suspicious activity and the ability to respond in such a way that would extricate them from a problematic situation.

Phase 1

Participants

Phase 1 of the multi-phase SAP development and piloting was conducted using four postgraduate psychology students. After receiving feedback, the stimuli were amended so that sections of the dialogue were clearer. Based on those responses, the original question, "what would you do if you were X and why?", was separated to ensure that both aspects (i.e., what *and* why) would be addressed at every segment. The concluding section of each scenario was also altered so that the relevant crimes were not revealed within the dialogue. This was to reduce the expectation for criminal activity in future scenarios and to gauge participants' knowledge of the crimes via further questioning. Subsequently, 10 participants were recruited from Amazon Mechanical Turk (MTurk)[1] to pilot the task in its online format and to provide data to assist in developing the coding protocol. Participants were required to be 18 years or older, live in Australia, New Zealand, United Kingdom, Canada, or the United States and speak English as their primary language to be eligible for the study. They were also required to hold "Master" status on CloudResearch (formerly TurkPrime; Litman et al., 2017) (i.e., they had adequately completed a substantial number of studies on the platform). This setting is useful for improving data quality by reducing the likelihood that bots or

[1] The reliance on online samples in this and subsequent studies was necessitated by (a) the fact that thesis data collection coincided with Covid-related restrictions and (b) the subsequent reluctance of potential participants to attend 'live' testing sessions at that time.

unreliable responders participate in the study. Responses from the first 10 participants were judged as sensible with no exclusions required, so a further 40 participants with the same requirements were recruited to create a total of 50 responses (21 males, 29 females; age range 27–69, $M = 43.88$, $SD = 9.88$). Again, no exclusions were required from this sample. The study was approved by the Flinders University Social and Behavioural Research Ethics Committee.

Materials

Suspicious Activity Paradigm (SAP). The original SAP comprised seven audio scenarios. Four scenarios depicted an unfolding crime while three scenarios were neutral (i.e., no crime was involved). Each crime scenario was designed to become progressively more suspicious as the story unfolded but maintained some ambiguity so as not to reveal criminal activity until the final segment when authorities or, in one case, a manager, became involved. In all crime scenarios, the fictional protagonist within the scenario was unaware of their potential involvement in criminal activity until the final segment when authorities became involved. The insinuated crimes included possession of child exploitation material, burglary, theft, and visa fraud. The three neutral, or non-crime, scenarios were designed to reduce the likelihood that participants would be primed to expect criminal activity in every scenario. In addition, the initial segment of all seven scenarios was designed to avoid arousing suspicion, thereby exposing participants who simply responded suspiciously to all cues. Three multiple choice questions that related to the scenario content were also included at the conclusion of each non-crime scenario as attention checks. Participants were allowed to continue with the task if any of the three attention checks were failed; however, responses were later reviewed for potential exclusions. The duration of each scenario's audio recording ranged from 111 to 238 seconds. The total time taken to complete each scenario involved both the recording and

participants' subsequent response times. After providing informed consent and basic demographic information, participants viewed the following instruction:

> "You are about to listen to several narrated scenarios. We want you to pay close attention, as in each scenario you will be asked to assume the position of one specific character. At certain points throughout the scenario, you will then be asked what you would do if you were in that character's position, and why. Please do not spend long on each answer - try to make a decision as quickly as you would in real life.
>
> The story may not always follow the choices you make, but please try and imagine what you would do if you were to find yourself in that position."

Each scenario was then divided into segments (ranging between 4 and 6), and after each segment participants were asked "What would you do if you were (character name) in this situation?" followed by "Please explain why you would do that". Both questions required open-ended responses. At each interval, participants were also required to indicate how sure they were of their responses on a four-point scale ranging from "Very unsure" to "Certain". This was included to allow for investigation of whether SAP responding was related to participants' confidence.

As each scenario was named after the story's protagonist, crime scenarios are henceforth referred to as Aaron, Andy, Amy, and Jia, and non-crime scenarios are referred to as Casey, Derek, and Charlie. Figure 1 displays the text version of the 'Aaron' crime scenario. The text version of all other scenarios appears in Appendix A (pp. 156-166). At the conclusion of the task, participants were asked what they believed happened at the end of each crime scenario. This was to differentiate understanding of the law from the inability to detect suspicious cues prior to criminal activity. The median task completion time for the original task was 64 minutes.

Segment 1:

One afternoon, Aaron is working his casual job in retail when his boss posts the roster for the coming fortnight. He sees that he has not been allocated many shifts and goes to speak to his manager.

> **Manager:** Sorry Aaron, I really don't have anything available for you at the moment because of the COVID pandemic. Everyone's hours are being cut.

Segment 2:

Aaron arrives home and looks at the unpaid bills piling up in his kitchen. He heads to his room to play video games, but the internet is running slowly because it hasn't been paid for. He becomes frustrated and starts looking for a second job on SEEK. As he is searching, Aaron's friend Jim comes over and sees that he is on SEEK.

> **Jim:** Hey, are you looking for another job?
>
> **Aaron:** Yeah, shifts have been cut back at work so I'm struggling to get money together for rent and bills. The internet has slowed because I'm late with a payment so I can't play Fortnite either.
>
> **Jim:** Well, if you keep some files on your computer for me, I'll give you some cash to help with internet and electricity, so you can still game. I've run out of space on my hard drive, so if you store some of my files it'll free up my computer without me needing to purchase more storage.

Segment 3:

> **Aaron:** Sure, thanks so much. That will really help me out.
>
> **Jim:** No problem, you're doing me a favour. The only thing is that you can't open my files, share them, or tell anyone anything about them.

Segment 4:

The next day, Jim brings over a USB with the files he wants stored and Aaron begins transferring the files to his computer. He saves them to cloud storage because he has plenty available, and it will keep the space on his own computer free. While the images are transferring, the image thumbnails show photos of young children.

Segment 5:

Jim notices Aaron looking at the thumbnails.

> **Jim:** Oh, don't worry about those, I've been studying photography and they are models for an assignment. It's all a surprise for my wife so please don't tell her about the course either.

Segment 6:

Aaron continues saving the images. Once finished, Jim thanks him and leaves.

One week later, Aaron hears a knock on the door while he is gaming. He is annoyed he has been interrupted but goes to answer the door. When the door opens, he sees two police officers.

Figure 1. *The Aaron crime scenario. Between each segment, participants are asked "What would you do if you were Aaron in this situation?" followed by "Please explain why you would do that". They then provide a rating of how sure they are of their response.*

Procedure

The study was presented on Qualtrics, an online survey platform (Qualtrics, Provo, UT). Participants first provided basic demographic information including age, gender, ethnicity and occupation. They also completed an initial attention check which required them to select the response option second from the right. If this attention check was failed, participants were redirected to a page that informed them they could not continue with the survey. They were also redirected to the study termination page if they attempted to complete the study on a mobile device. They were then provided with the task instructions and began the task. All scenarios were presented in random order.

Coding protocol. The initial piloting data (*N*=50) were used to develop a coding protocol that was likely to capture most participant responses at each segment. Categorical codes for the qualitative responses included both a behavioural aspect (e.g., "I would not do what they ask") and recognition of suspicious cues (e.g., "because I think they are asking me to do something dodgy/illegal/wrong"). For example, participants would accept an offer/comply with a request, decline an offer/deny a request, respond conversationally (e.g., I would smile and ask how their day has been"), seek clarification or further information, extricate themselves, or involve authorities. An additional code was later included to capture non-behavioural responses (e.g., "I would feel scared"). All responses that were not readily captured by the codes above were coded as 'Other'. Coding for the suspicion scale included making no reference to suspicion, indecisively referencing suspicion (e.g., being curious or finding something "weird"), being suspicious, and stating that the situation is not suspicious. The initial protocol, developed using responses from the first 50 participants, included codes, definitions and several examples, and was later expanded to include multiple examples for each scenario as additional data were collected, as well as refined definitions. Figure 2 displays the first version of the coding protocol which was used to conduct preliminary inter-

rater reliability analyses, while the complete coding protocol with updated response examples and definitions appears in Appendix B (pp. 167-169). Coding examples for one scenario (Aaron) are displayed in Figure 3.

BEHAVIOUR CODES *(What would you do if you were X in this situation?)*

(1) Comply with request/accept offer/go ahead
- The participant agrees when they are asked to do something by another character
- The participant indicates that they will take up an offer from another character
- The participant decides to 'go ahead' and continue down the path of the story. Generally relevant in the segments after an offer or request has been made, or when no other character is involved

(2) Take action
- The participant indicates that they would do something other than continue along the path of the story, such as involving another character. It does not necessarily remove them from the situation.

(3) Deny request/decide against/decline offer
- The participant indicates that they will not do what is asked of them by another character
- The participant indicates that they would turn down an offer made to them
- The participant decides not to continue with the path of the story (when no offer/request is made)

(4) Extricate
- The participant indicates they would do something which actively removes them from the situation. This does <u>not</u> include contacting authorities but does include involving a manager or supervisor.

(5) Conversational/socially acceptable
- The participant responds in a way that would be considered conversational or socially acceptable. This might be relevant during conversations or when participants suggest they would act in a way that would be considered polite or socially motivated.

(6) Seek clarification/further information
- The participant indicates that they would seek further information, or would like something clarified, before making a decision.
- This is relevant any time they would ask a question that is important to the progression of the story. If the question is just conversational, code as a 5.
- Does not always need to involve asking a question of somebody - can include doing research to find more information.

(7) Continue as were/no action taken
- The participant indicates that they would not take any real action in that situation or would continue with the same activity as before.

(8) Involve authorities
- The participant indicates that they would call authorities (such as police). This does not include making contact with a manager or supervisor.

(9) Non-behavioural response
- The participant responds in a way that does not denote a behaviour but rather states how they would feel about the situation (e.g. "I would feel nervous").

(10) Other/irrelevant
- Any response that does not readily fit into these categories.

SUSPICION CODES *(Why would you do that?)*

(0) No reference to suspicion
- The participant provides an explanation that has no reference to suspicion or is irrelevant.

(1) Reference to suspicion (indecisive)
- The participant expresses some hesitancy but does not suggest they are suspicious.
- Curiosity could be coded here.
- E.g. “there could be multiple explanations for this behaviour” or “not enough details about the job to assess its validity yet”.

(2) Suspicious
- The participant expresses some suspicions about the situation. Being suspicious is defined as:
  - having or showing a cautious distrust of someone or something
  - causing one to have the idea or impression that someone or something is questionable, dishonest, or dangerous
  - having the belief or impression that someone is involved in an illegal or dishonest activity.
- The participant expresses doubt, dubiety, mistrust, scepticism, or uncertainty about the situation.

(3) Not suspicious
- The participant references suspicion but states they believe the situation is not suspicious, or that they have no reason to be suspicious at that time. Responses that include trusting the other character would be coded here.

If participant refers to suspicion in the ‘behaviour’ section (i.e. *What would you do?),* still count this and code under the ‘explanation’ section. Similarly, if the participant gives an ambiguous response in the ‘behaviour’ section that is given context in the ‘explanation’ section, take this into account.

Figure 2. *Version 2 of the coding protocol*

AARON BEHAVIOUR CODING EXAMPLES

(1) Comply with request/accept offer/go ahead
- “I would agree to store the files”
- “I would accept the offer"
- “I would agree, do not really want to know what is in the files anyway”
- “I would be open to storing the files”

(2) Take action
- **Segment 1: Any response that involves looking for another job**
- “start looking for another job”
- “I would find out if there were covid related unemployment benefits available and apply for them.”
- “tell him that for legal reasons I would want him to put that in writing and sign it and agree to be recorded”
- “I would confirm what the images are”
- “I would feel compelled to look”

- "I would tell Jim that I understand but that seeing the photos made me uncomfortable. I would explain that I have no interest in his private matters but I wasn't expecting to see thumbnails of nothing but children."
- "I'd confront Jim's lies."

(3) Deny request/decide against/decline offer

- "I would tell him that's okay, I should probably be focusing on getting another job anyways and not playing video games."
- "I would say no to the offer"
- "I would tell him never mind, I'd rather not have the files on my computer, that seems like a big responsibility."
- "I would still be uncomfortable and tell him he needs to find other storage."
- "I would stop downloading and ask my friend what the heck?"

(4) Extricate

- **(relevant after the downloads have begun. Must include action to remove themselves from the situation (i.e. deleting the files), not just stopping the download)**
- "I would stop loading the files and delete the others"
- Stop transferring the files, delete whatever I managed to upload to my cloud storage and tell my friend the deal is off."

(5) Conversational/socially acceptable

- **Segment 1: Any response that involves discussing hours or employment opportunities with the boss would be coded here.**
- "I would say that I am available whenever possible."
- "I would explain my situation and ask if there's any way I could pick up some additional work."
- "I would try and see if there are really no extra shifts left."
- "I would explain to her that I understand the situation and ask her to notify me when and if she has more hours available."

(6) Seek clarification/further information

- "I would ask why hours were being cut because of the COVID pandemic."
- "find out what kind of files they are"
- "I would want to make sure that the files I am getting are legal"
- "I would ask to make sure they weren't porn files or something else illegal"
- "Be curious why there is young children may ask about this"

(7) Continue as were/take no action

- "I know the man so I would take what he says at face value"
- "This sounds genuine and if the kids are clothed and happy looking I would believe it"

(8) Involve authorities

- "Call the cops immediately"
- "I would immediately contact the authorities and hand over the hard drive to them, while also giving them the information on who I got the hard drive from and where they can be located."
- "I would tell him that it doesn't make sense and ask him for more information. During this exchange I would be recording him secretly so that I could then take all the information I have to the authorities."

(9) Non-behavioural

- "I would be grateful."
- "I would be disappointed but I know it's not her fault."

(10) Other

- "It depends, are the children in a negative image like child porn or are they just images that could be for a catalogue or family photos"

AARON SUSPICION CODING EXAMPLES

(0) No reference to suspicion
- "I don't want to blame someone else for something that's out of their control"
- "Because I can't pay the bills on the reduced hours."
- "I need more money and need to be responsible."
- "I believe in keeping people's private things private. I would have no interest or need to open up the files, nor to share them. I wouldn't betray my friend that way."

(1) Ref suspicion (indecisive)
- "There could be multiple explanations for his behaviour"
- "I would still be sceptical, but as long as the children look normally dressed and proper, I guess I would believe my friend's story and go ahead and store the files."
- "I would need to see damaging photos to turn someone in, it could ruin their life if I am unsure"

(2) Suspicious
- "Seeing young children seems odd, it does not feel right".
- "It sounds fishy to me"
- "I don't want any illegal files on my computer"
- "I would be immediately suspicious about the types of files I would be storing and would want to make sure that the files I am getting are legal."
- "I'd be concerned about accidentally storing something illegal. It's pretty suspicious to offer to pay someone to store files, since storage is so cheap."
- "Him telling me I can't see them or tell anyone about them makes me nervous about what they are and if it may be something illegal."
- "I don't want people to think that I had anything to do with such a thing."
- "I'm worried that my buddy might be a pedophile"

(3) Not suspicious
- "At this point, I have no reason to think she isn't telling me the truth so this seems like the best course of action."
- "Since they are my friend, I know they wouldn't have me store anything inappropriate."
- "He's my friend so I trust him."

Figure 3. *Example responses from Crime scenario 1 (Aaron)*

Results

Inter-rater Reliability

For the initial assessment of inter-rater reliability for the coding protocol, a second post-graduate psychology student familiar with the project independently coded the first 10 participants' responses using the coding protocol displayed in Appendix B (pp. 167-169). Prior to independent coding, a training session was conducted to discuss general coding guidelines and definitions. After the responses were coded, I met with the second coder to

discuss discrepancies. Initially, it was deemed appropriate to select multiple codes when relevant to capture as much detail as possible. For example, in reaction to segment 4 of the Aaron scenario, the response "I would tell him that it doesn't make sense and ask him for more information. During this exchange I would be recording him secretly so that I could then take all the information I have to the authorities", could have been coded as "seek further information" and "contact authorities". Consequently, there were multiple occasions on which two codes were assigned but only one code was agreed upon. Inter-rater reliability (Cohen's kappa; McHugh, 2012) was first calculated by considering at least one common code as indicating agreement. Reliability for the behaviour responses lay between k = .36 and k = .86, and for the suspicion responses between k = .54 and k = .79. As a result, the reliability of several scenarios was considered unacceptably low: see Table 1 for the reliability coefficients.

Table 1.

First round of inter-rater reliability for coding of each scenario.

Scenario	Behaviour		Suspicion	
	% Agreement	Kappa	% Agreement	Kappa
Aaron	88.0	.86	80.0	.68
Andy	60.0	.64	55.0	.54
Amy	44.0	.36	82.0	.73
Jia	78.0	.72	88.0	.79
Casey	52.0	.40	86.0	.73
Derek	85.0	.80	85.0	.64
Charlie	64.0	.56	74.0	.61

Coding discrepancies were reviewed, and agreement was generally reached; however, discussion around coding disagreements highlighted a need for further clarity in the coding protocol. Subsequently, examples from participants' responses were included under each relevant code in the protocol (see Figure 3 for examples), and a clear definition of suspicion

was included. Suspicion was defined as (1) having or showing a cautious distrust of someone or something, (2) causing one to have the idea or impression that someone or something is questionable, dishonest, or dangerous, or (3) having the belief or impression that someone is involved in an illegal or dishonest activity. Once additional data were collected (*n*=40), the second coder independently coded a random selection of 10 additional participants. Inter-rater reliability values calculated from the second round of coding are shown in Table 2. Again, at least one common code was considered to be agreement in the analyses. The process of discussing discrepancies was repeated and further refinements made to the protocol. Emphasis was placed upon only one code being assigned to a response where possible, and the non-crime scenario, 'Charlie', was removed due to difficulty coding and poor inter-rater reliability. Furthermore, due to the heterogeneity of responding, it was decided that any responses difficult to categorise would be discussed between coders in the future as a further safeguard to improve coding reliability.

Table 2.

Second round of inter-rater reliability for coding on each scenario.

Scenario	Behaviour		Suspicion	
	% Agreement	Kappa	% Agreement	Kappa
Aaron	86.0	.83	86.0	.77
Andy	88.3	.85	83.1	.75
Amy	92.0	.89	82.0	.74
Jia	82.0	.76	86.0	.76
Casey	78.0	.72	68.0	.44
Derek	92.5	.89	92.5	.83
Charlie	68.0	.62	64.0	.45

Confidence Ratings

To gauge the level of confidence with which individuals responded to suspicious cues, participants were asked the follow-up question "How sure are you about that?" after the

behaviour and suspicion questions of each segment. During initial piloting, this scale was presented with four anchor points ranging from “Very unsure” to “Certain”. Review of the pilot data revealed that “Very sure” and “Certain” were endorsed 96.7% of the time (N=1750). Perusal of the qualitative responses to the initial questions (i.e., what would you do and why?) revealed a common discrepancy between language that would suggest sureness and responses to the ‘sureness’ scale (i.e., participants used language such as “I’m pretty sure” but indicated they were “certain”). As a result, the scale was amended to present sureness on a ten-point scale ranging from 0% to 100%. It was hoped that the updated scale would reveal more variance in the confidence of decision-making.

Phase 2

Given the variability in responses from the initial pilot sample, the purpose of the second phase of piloting was to create a scoring protocol. Creation of a scoring protocol was complicated by the subjective nature of the stimuli content, whereby the objective classification of a response as either ‘appropriate’ or ‘inappropriate’ at any particular timepoint was difficult to determine. It was anticipated that there would be multiple ‘appropriate’ responses in some segments: for example, two people might find a situation suspicious but react in different ways (e.g., extricating or seeking further information), before reaching the same point (e.g., extrication) in a subsequent segment. Consequently, I attempted to develop guidelines as to what might reasonably be considered to be the appropriate response at each segment. To provide a snapshot of what might be considered the most appropriate responses, three separate samples were recruited to complete the task. First, a small group of post-graduate psychology students met to conduct a group discussion about what they believed to be the most appropriate responses for each step. Next, a sample of four legal professionals completed the task individually in its online format. This group was selected because their professional training was considered to have equipped them with the

ability to provide an indication of the most appropriate response to each scenario. Finally, a larger online sample was collected to provide a further perspective of what might be considered appropriate in a broader sample. It was anticipated that there would be relative consistency in the responding among these samples, with a pattern of responding that demonstrated a general increase in suspicion and appropriate behavioural responses that could be used as a comparison point for future studies.

Participants

Postgraduate Psychology Students. Eight post-graduate psychology students were recruited from Flinders University to complete the task as a collective. Participants listened to the audio clips aloud as a group and provided their individual response on Qualtrics via laptops or tablets. After each individual response had been recorded, the group convened as a panel to discuss any differences of opinion and collectively decide on the most appropriate responses for each segment. It was anticipated that each segment may have more than one appropriate response. For instance, in one situation it may have been deemed equally appropriate to decline an offer or seek further information, while in another it may have only been appropriate to decline the offer. This process took two hours.

Lawyers. Four individuals who were either previously or currently employed as a lawyer completed the study independently in its online format. This sample was targeted to provide a manipulation check for the SAP by indicating whether the proposed suspicious cues were detectable by individuals with professional training and expertise in legal matters and wrongful conduct.

Broader sample. An additional group of 115 online participants was sourced from Amazon Mechanical Turk to complete the task. Worker requirements remained the same, except for the ‘Master worker’ rule. Due to difficulty sourcing participants that met those criteria after excluding previous studies, the study was opened to any workers who met the

age, language, and nationality requirements. It was assumed that the quality of the data would not be compromised by this decision because the length and requirements of the study appeared to deter previous participants who were not willing to participate fully (Phase 1 had a 63% completion rate). Similarly, the large volume of open-ended questions allowed for non-sensical or inappropriate responding to be easily identified and excluded when necessary. One hundred and fifteen participants were targeted to allow for possible exclusions. Only two exclusions were required due to uninterpretable responses, leaving a sample of 113 participants (59 males, 54 females; age range 18–71, $M = 40.50$, $SD = 11.64$).

Measure

Three minor adjustments were made to the task at this time: (1) the non-crime scenario named ‘Charlie’ was removed, (2) the presentation of the certainty rating was changed from a four-point scale (‘Completely Unsure’ to ‘Certain’) to a ten-point scale (0% to 100%), and (3) the instructions were augmented slightly to increase clarity around the expectations for responding. A small subset of responses in the previous data indicated that participants felt the need to respond as they thought the character would, rather than how they themselves would react. For example, on the ‘Aaron’ crime scenario, one participant stated “I would wonder why I am stupid enough to believe Jim and why I am stupid enough to keep transferring these files. But since this survey continues to depend on Aaron doing stupid things, I suppose if I were Aaron, I would just shrug and keep transferring the files.” While the study instructions previously noted that the story was pre-determined and would therefore not update according to the participants’ responses, it was further clarified that answers should reflect what the participant would do in that situation, not what they believed the character would do. The median completion time for the task in its final form was 64 minutes.

Results

Postgraduate Psychology Students

Individual responses within the postgraduate psychology sample showed that recognition of suspicion differed across the group, despite the potential influence of the group discussion. Nevertheless, despite the variation in responding throughout the scenarios, by the final segment in each crime scenario the group had agreed that something suspicious had occurred. Debate about the most appropriate course of action revealed that several behaviours were considered to be appropriate, and discussion occasionally influenced responses to become dependent on additional context not included in the scenario (e.g., "If this happened, I would do X, but if that happened, I would do Y"). At times, the group was unable to reach a consensus which suggests that the SAP is not suitable for use in a group and is best administered individually, as originally intended. Therefore, subsequent testing was all conducted at an individual level. Appendix C (pp. 170-173) shows the individual responses from the postgraduate psychology sample.

Legal Sample

The sample of four current/prior lawyers, who completed the study independently, consistently recognised suspicious activity by the final segment of each scenario. While the participants recognised suspicion at varying stages, at least three out of the four participants indicated suspicion by the final segment of each crime scenario, demonstrating that the scenarios adequately depicted suspicious or 'dodgy' situations that warranted concern. Responses from the legal sample can be found in Appendix D (pp. 174-175).

Broader Sample

Inter-rater reliability. To ensure that the coding protocol was easily and reliably implemented by multiple coders, a second round of inter-rater reliability checks was conducted with a research assistant who was new to the project. The training process was

repeated; it consisted of an initial session describing the project, presentation of an example scenario, and review of the core definitions of each code. The responses from 10 participants were then coded independently and compared for discrepancies. This process was repeated twice more, until acceptable inter-rater reliability was reached (see Table 3 for final inter-rater statistics and Appendix E [p. 176] for rounds one and two). At this point, it was decided that a single code should be assigned to each response to increase reliability and to provide clarity in future statistical analyses.

Table 3.

Third round of inter-rater reliability for coding on each scenario.

Scenario	Behaviour		Suspicion	
	% Agreement	Kappa	% Agreement	Kappa
Aaron	90.0	.89	88.0	.79
Andy	90.8	.87	89.1	.79
Amy	86.0	.82	89.0	.81
Jia	81.0	.75	91.0	.80
Casey	82.0	.77	92.0	.86
Derek	80.0	.70	90.0	.73

‘Most appropriate’ responses. Patterns of responding for the broader sample varied (See Appendix F, pp. 177-178). When considering all 10 behaviour codes individually, it became clear that a standout ‘most appropriate’ response for the behaviour aspect of every segment would be difficult to determine. Therefore, the behaviour codes were condensed into three broader categories, with the intention of retaining the specificity of the codes while allowing for clearer patterns of responding to emerge. The three resulting categories were:

1. The person took steps to extricate themselves from the situation or chose to act in a way that would interrupt the progression of the story (e.g., take action, deny/decline, extricate, or contact authorities).

2. The person did not act in a way that removed them from the situation (e.g., comply/accept, continue as were, conversational).

3. The person sought further information or clarification before deciding how to proceed.

Responses classified under the 'non-behavioural' and 'other' codes, as well as missing data due to audio failures made up 5.37% of behaviour responses across all scenarios (total responses*:* $N = 3390$) and were excluded from analyses when codes were combined. To determine whether there was a significant difference between the percentage of responses that fell under each code at each step of the story, Chi-Square Goodness of Fit tests were conducted for each segment in each scenario. Due to the absence of previous research upon which population estimates could be derived, analyses were run with the assumption that all categories had equal probability of representation. Post-hoc binomial tests with Bonferroni correction were then conducted to determine whether a single behaviour could be considered the 'most appropriate' at each segment by identifying statistically significant differences between the code with the highest percentage of responses and the remaining codes. Due to the variation across scenarios, these are discussed individually.

Crime scenario 1 (Aaron). Figures 4 and 5 show the patterns of responding across the scenario. Chi-Square Goodness of Fit tests (see Table 4) showed that the percentages of each response differed from the population estimate (33.33%) in all segments. Post-hoc analyses revealed that by segment 4 the majority of participants provided an extricating behavioural response, with a significant difference in percentage between those who extricated themselves (82%) and those who sought further information (18%), $p<.001$, Cohen's g^2 = 0.32. This pattern of responding was repeated, albeit to a lesser degree, in segment 5

[2] Cohen's *g* sourced from Cohen (1988, pg. 147). Cohen's interpretation of this effect size is noted as: 0.00-<0.05 = negligible, .05-<0.15 = small, 0.15-<.25= medium, and ≥0.25 = large.

(extricate = 67%, seek info = 33%, p = .008, g = .17). Tables 1 and 2 in Appendix G (pp. 179-180) contain all pairwise comparisons. The pattern of suspicion recognition was similar to that of the behavioural responses. In all segments, the percentage of responses within each code differed from the population estimate (25%). Suspicion was recognised by the majority of participants in segment 3 (suspicion = 65%, no reference = 35%, p = .015, g = .15) and remained the most prevalent category thereafter (segment 4: suspicion = 90%, indecisive = 10%, p<.001, g = .40; segment 5: suspicion = 85%, indecisive = 15%, p<.001, g = .35).

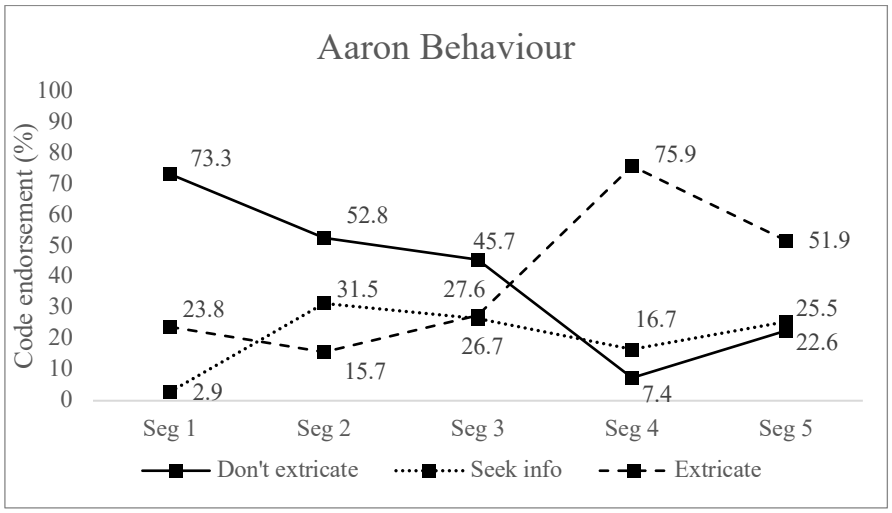


Figure 4. *Line graph representing the percentage of each behaviour code at each segment in the Aaron scenario*.

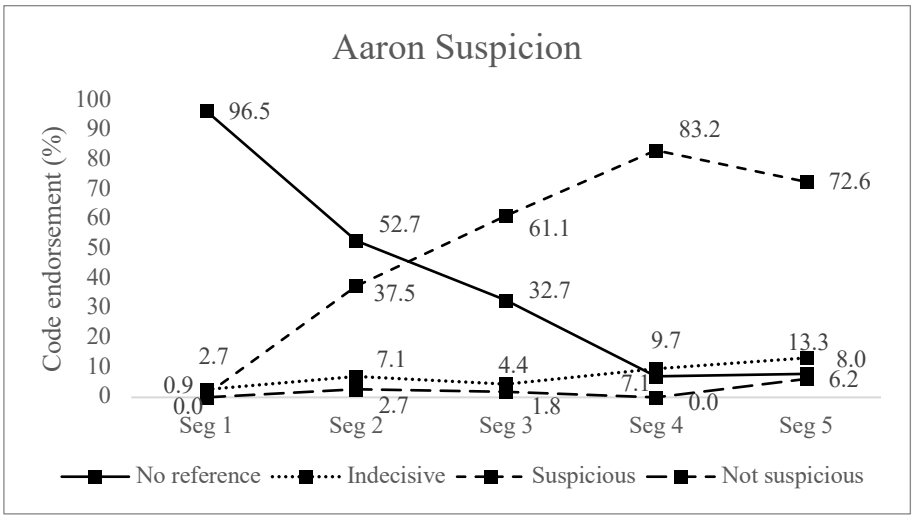


Figure 5. *Line graph representing the percentage of each suspicion code at each segment in the Aaron scenario*.

Table 4.
Chi-square Goodness of Fit tests for the Aaron scenario.

	One Sample Chi Square	Cramer's V^3
Behaviour		
Segment 1	χ^2 (2) = 82.51, p <.001	0.63
Segment 2	χ^2 (2) = 22.39, p <.001	0.32
Segment 3	χ^2 (2) = 7.26, p =.027	0.19
Segment 4	χ^2 (2) = 88.56, p <.001	0.64
Segment 5	χ^2 (2) = 16.55, p <.001	0.28
Suspicion		
Segment 1	χ^2 (2) = 202.69, p <.001	0.95
Segment 2	χ^2 (3) = 77.93, p <.001	0.48
Segment 3	χ^2 (3) = 105.02, p <.001	0.56
Segment 4	χ^2 (2) = 126.50, p <.001	0.75
Segment 5	χ^2 (3) = 137.58, p <.001	0.64

Crime scenario 2 (Andy). Goodness of Fit tests confirmed that percentages of each code differed from the population estimate in all segments (see Table 5). Descriptive statistics (see Figures 6 and 7) demonstrated that at no point throughout the scenario did the majority of participants decide to extricate themselves. Post-hoc binomial tests revealed that only segments 4 and 6 produced a behaviour response that was significantly different from all others, with participants choosing not to extricate themselves more often than their choosing of both other behaviour categories (segment 4: don't extricate = 89%, seek info = 11%, p<.001, g = .39; segment 6: don't extricate = 76%, seek info = 24%, p<.001, g = .26). In segments 1 to 3, there was no statistically significant difference between seeking information and not extricating oneself, indicating that the scenario was likely not suspicious enough to warrant sufficient concern (see Table 3 in Appendix G [p. 181] for all pairwise comparisons). This was reflected in the suspicion responses, where no reference to suspicion was made in the majority of cases until the fifth segment, where being suspicious became the most

[3] The benchmarks for interpretation of Cramer's V when df = 3 are suggested as: 0-.06 = negligible, >.06-.17 = small, >.17-.29 = moderate, >.29= large (Cohen, 1988)

common response by a significant margin (suspicion = 65%, no reference = 35%, $p = .028$, $g = .15$).

Table 5.
Chi-square Goodness of Fit tests for the Andy scenario.

	One Sample Chi Square	Cramer's *V*
Behaviour		
Segment 1	χ^2 (2) = 58.62, p<.001	0.52
Segment 2	χ^2 (2) = 50.44, p<.001	0.48
Segment 3	χ^2 (2) = 24.60, p<.001	0.34
Segment 4	χ^2 (2) = 127.62, p<.001	0.76
Segment 5	χ^2 (2) = 8.65, p=.013	0.20
Segment 6	χ^2 (2) = 36.35, p<.001	0.41
Suspicion		
Segment 1	χ^2 (3) = 171.27, p<.001	0.72
Segment 2	χ^2 (3) = 98.55, p<.001	0.54
Segment 3	χ^2 (3) = 74.07, p<.001	0.47
Segment 4	χ^2 (3) = 92.84, p<.001	0.52
Segment 5	χ^2 (3) = 58.50, p<.001	0.42
Segment 6	χ^2 (3) = 74.07, p<.001	0.47

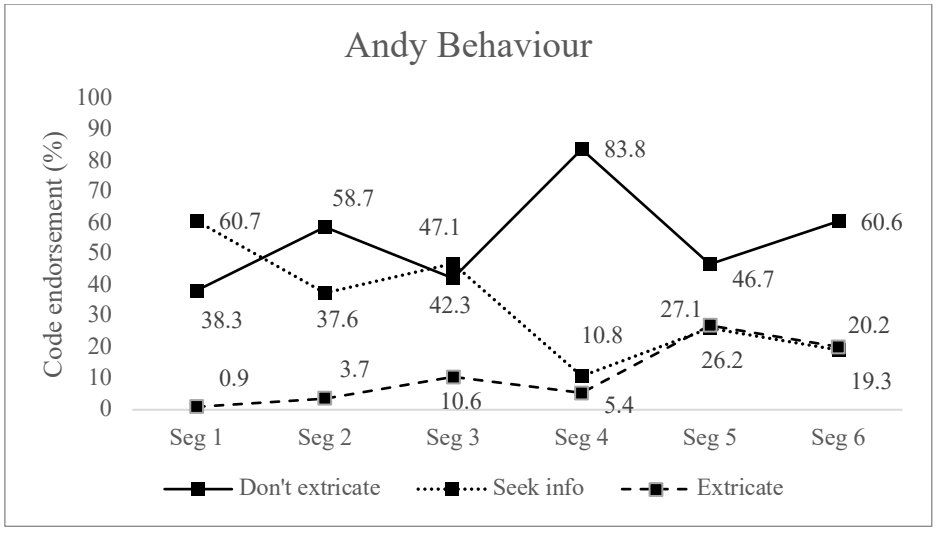


Figure 6. *Line graph representing the percentage of each behaviour code at each segment in the Andy scenario*.

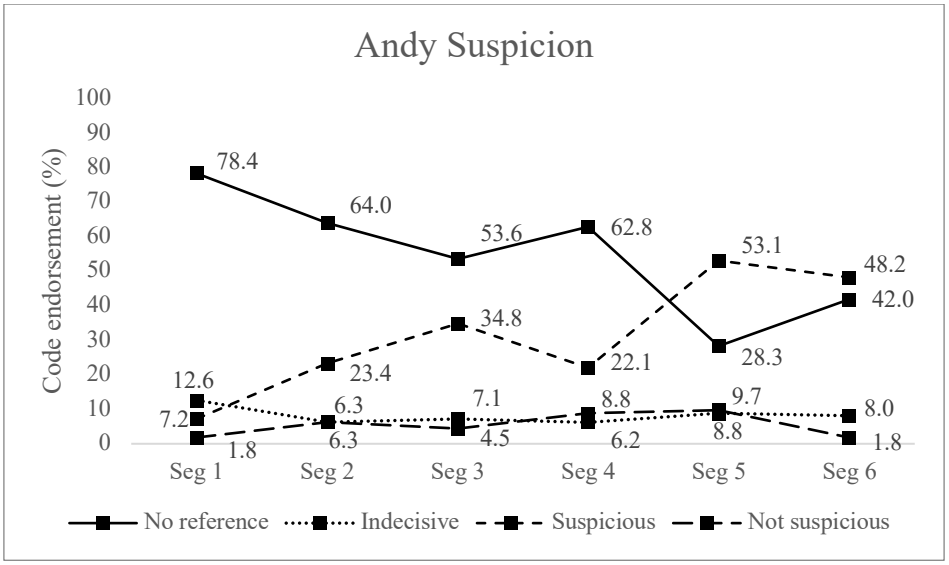


Figure 7. *Line graph representing the percentage of each suspicion code at each segment in the Andy scenario*.

Crime scenario 3 (Amy). Behavioural responses on the Amy scenario also differed significantly from one another in every segment (see Table 6). Descriptive statistics (see Figure 8) revealed a dramatic pattern of responding with extricating oneself the most common behavioural response from segment 2 onwards. This was the preferred response by a considerable margin in segments 2 (extricate = 97%, don't extricate = 3%, $p<.001$), 3 (extricate = 82%, don't extricate = 18%, $p<.001$) and 5 (extricate = 97%, don't extricate = 3%, $p<.001$). Although extricating appeared to be the preferred response in segment 4 (see Figure 9), the difference between extrication and non-extrication was statistically non-significant. In comparison to the behaviour responses, recognition of suspicion was delayed, indicating that extricating behaviours were considered the appropriate course of action independent of any perceived malicious intent from the antagonist. Suspicion was not significantly represented until the final segment (suspicious = 94%, indecisive = 6%, $p<.001$), at which point, the difference was large (g = .44).

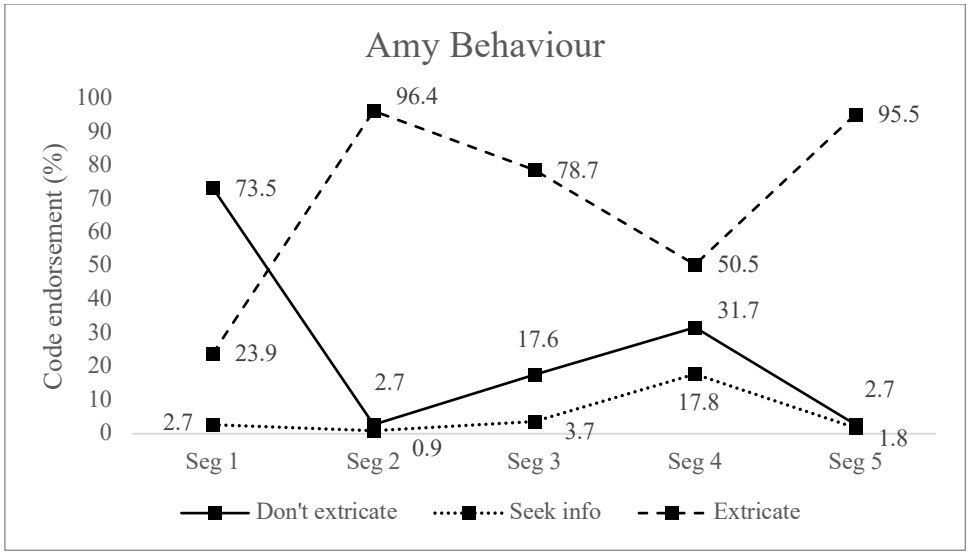


Figure 8. *Line graph representing the percentage of each behaviour code at each segment in the Amy scenario*.

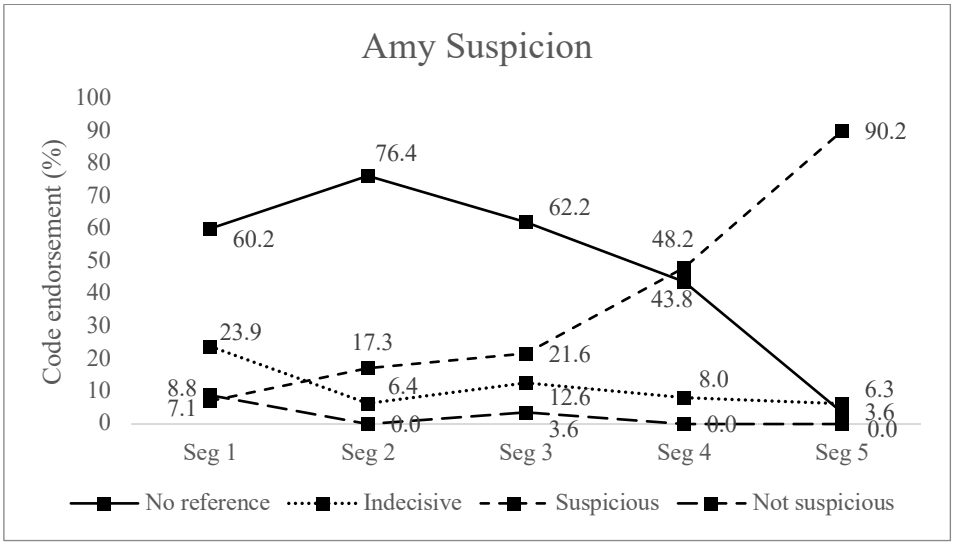


Figure 9. *Line graph representing the percentage of each suspicion code at each segment in the Amy scenario*.

Crime scenario 4 (Jia). Behaviour responses in the Jia scenario deviated from one another in all except the second segment where there was an even distribution between seeking further information (54%) and non-extricating behaviours (46%) (see Table 7). Descriptive statistics (see Figures 10 and 11) and post-hoc analyses revealed statistically significant preferences for non-extricating behaviours at segments 1 (don't extricate = 93%, seek info = 7%, $p<.001$, $g=.43$) and 3 (don't extricate = 80%, seek info = 20%, $p<.001$, $g=$

.30) and extricating behaviours at segment 4 (extricate = 66%, seek info =34%, p = .006, g = .16). In comparison, suspicion was not recognised by the majority of participants until the final segment (suspicious = 66%, no reference = 34%, p = .010, g = .16).

Table 6.
Chi-square Goodness of Fit tests for the Amy scenario.

	One Sample Chi Square	Cramer's V
Behaviour		
Segment 1	χ^2 (2) = 89.49, p <.001	0.63
Segment 2	χ^2 (2) = 196.71, p <.001	0.95
Segment 3	χ^2 (2) = 103.17, p <.001	0.69
Segment 4	χ^2 (2) = 16.30, p <.001	0.28
Segment 5	χ^2 (2) = 191.04, p <.001	0.93
Suspicion		
Segment 1	χ^2 (3) = 82.29, p <.001	0.49
Segment 2	χ^2 (2) = 93.62, p <.001	0.65
Segment 3	χ^2 (3) = 88.96, p <.001	0.52
Segment 4	χ^2 (2) = 32.59, p <.001	0.38
Segment 5	χ^2 (2) = 162.98, p <.001	0.85

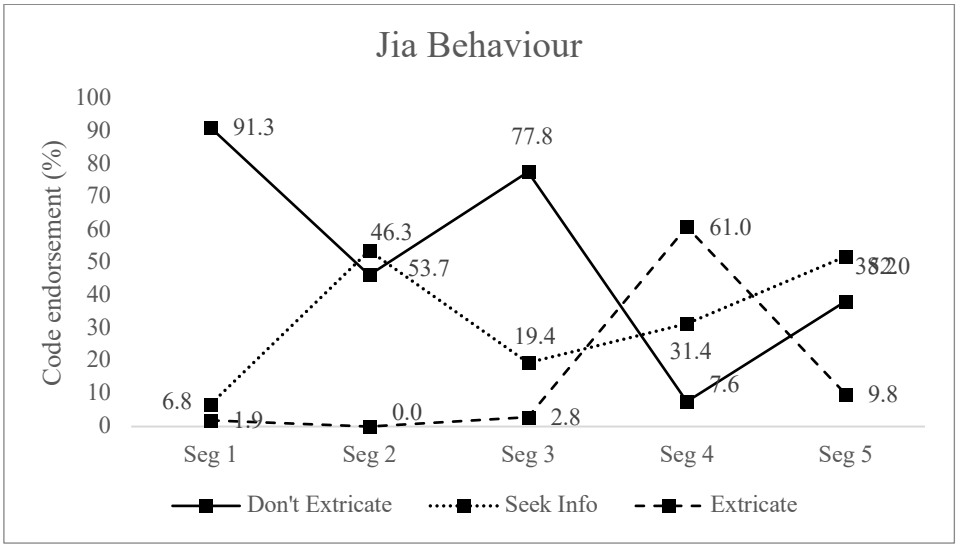


Figure 10. *Line graph representing the percentage of each behaviour code at each segment in the Jia scenario*.

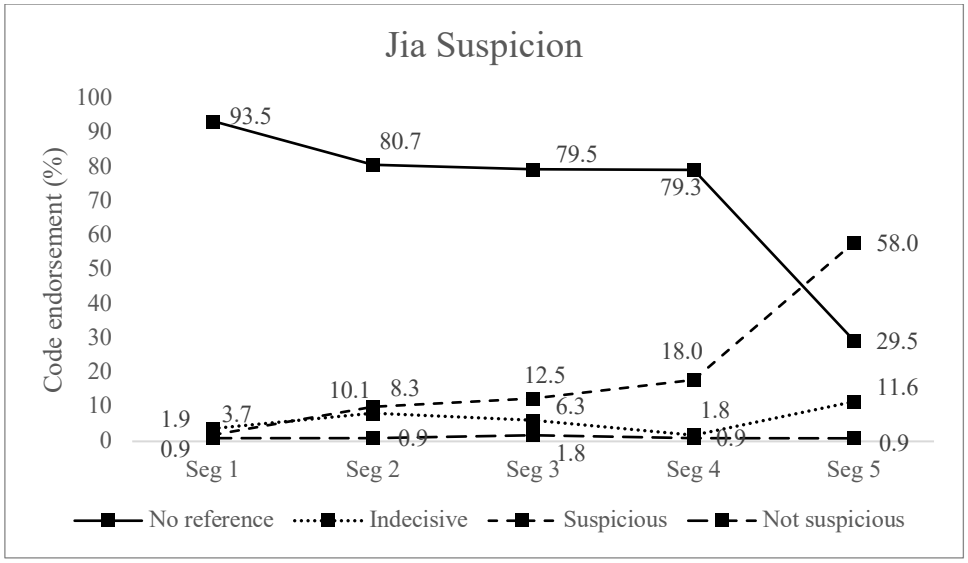


Figure 11. *Line graph representing the percentage of each suspicion code at each segment in the Jia scenario*.

Table 7.
Chi-square Goodness of Fit tests for the Jia scenario.

	One Sample Chi Square	Cramer's *V*
Behaviour		
Segment 1	χ^2 (2) = 155.90, p <.001	0.87
Segment 2	χ^2 (1) = 0.59, p =.441	0.07
Segment 3	χ^2 (2) = 100.50, p <.001	0.68
Segment 4	χ^2 (2) = 44.97, p <.001	0.46
Segment 5	χ^2 (2) = 28.29, p <.001	0.37
Suspicion		
Segment 1	χ^2 (3) = 270.59, p <.001	0.91
Segment 2	χ^2 (3) = 182.63, p <.001	0.75
Segment 3	χ^2 (3) = 179.79, p <.001	0.73
Segment 4	χ^2 (3) = 182.66, p <.001	0.74
Segment 5	χ^2 (3) = 83.86, p <.001	0.50

Non-crime scenario 1 (Casey). Behaviour responses on the Casey scenario also differed significantly from one another in every segment (see Table 8). As anticipated, the Casey scenario primarily elicited non-extricating behaviours (see Figure 12). Despite this, suspicion became the most endorsed code from the fourth segment onwards (see Figure 13), and by a significant margin in the final segment (suspicious = 66%, indecisive = 34%, *p*

= .006, *g* = .16). This was likely due to the final segment containing a less ambiguous cue than those that preceded it, with the protagonist, Casey, directly (although mistakenly) accused of possessing her friend's stolen bike.

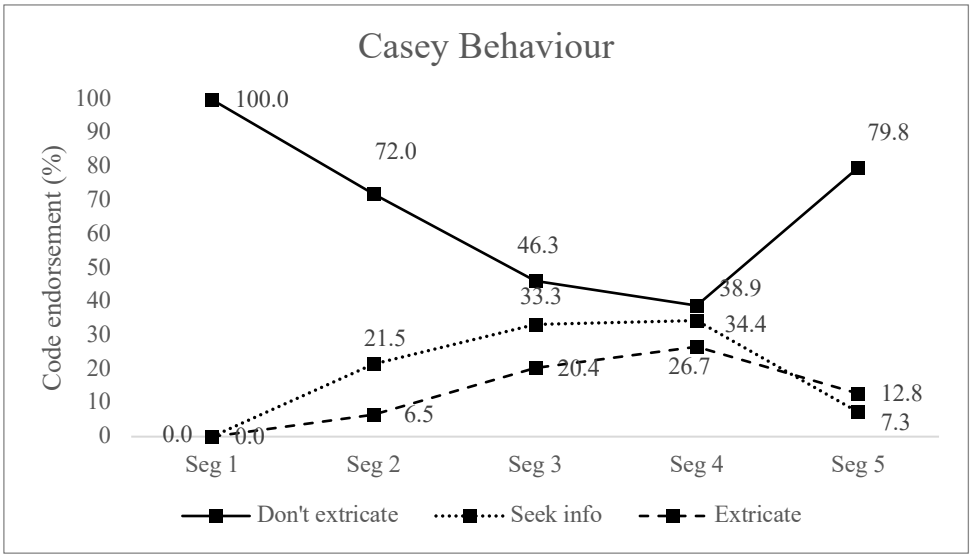


Figure 12. *Line graph representing the percentage of each behaviour code at each segment in the Casey scenario*.

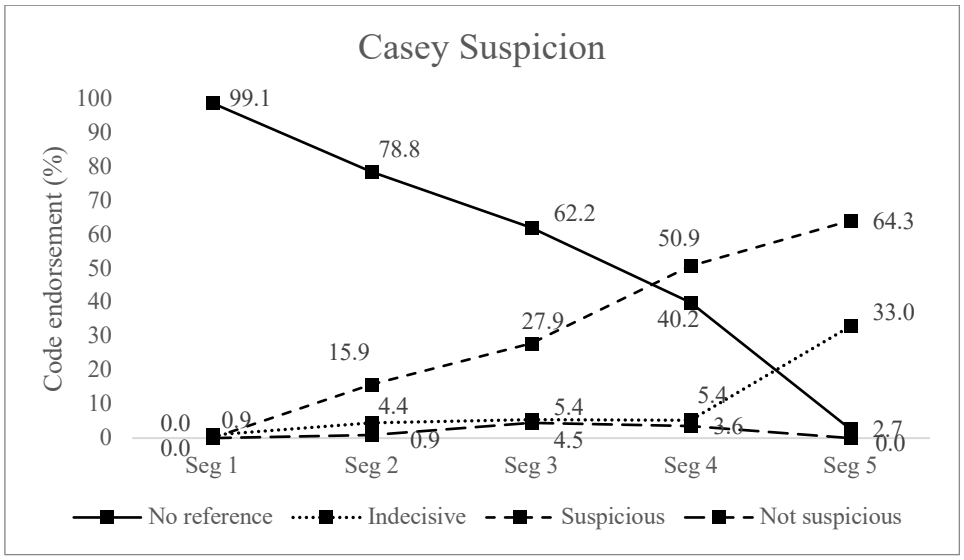


Figure 13. *Line graph representing the percentage of each suspicion code at each segment in the Casey scenario*.

Table 8.
Chi-square Goodness of Fit tests for the Casey scenario.

	One Sample Chi Square	Cramer's *V*
Behaviour		
Segment 1	-	
Segment 2	χ^2 (2) = 75.44, p <.001	0.59
Segment 3	χ^2 (2) = 10.89, p =.004	0.22
Segment 4	χ^2 (2) = 2.07, p =.356	0.11
Segment 5	χ^2 (2) = 106.48, p <.001	0.70
Suspicion		
Segment 1	χ^2 (1) = 107.04, p <.001	0.98
Segment 2	χ^2 (2) = 179.78, p <.001	0.89
Segment 3	χ^2 (3) = 97.40, p <.001	0.54
Segment 4	χ^2 (3) = 78.21, p <.001	0.48
Segment 5	χ^2 (2) = 63.77, p <.001	0.53

Non-crime scenario 2 (Derek). The final non-crime scenario, in which the protagonist, Derek, is asked to store some of his friend's belongings while she is away on holiday, demonstrated that participants had not simply been primed to suspect criminal activity in every scenario. Responses differed from the population estimate in both the behavioural and suspicion responses (see Table 9).

Table 9.
Chi-square Goodness of Fit tests for the Derek scenario.

	One Sample Chi Square	Cramer's *V*
Behaviour		
Segment 1	χ^2 (2) = 37.84, p <.001	0.41
Segment 2	χ^2 (2) = 71.70, p <.001	0.56
Segment 3	χ^2 (2) = 114.22, p <.001	0.72
Segment 4	χ^2 (2) = 68.05, p <.001	0.55
Suspicion		
Segment 1	χ^2 (3) = 109.29, p <.001	0.57
Segment 2	χ^2 (3) = 172.49, p <.001	0.71
Segment 3	χ^2 (3) = 105.64, p <.001	0.56
Segment 4	χ^2 (1) = 63.94, p <.001	0.75

Non-extricating behaviours were demonstrated by the majority of participants in all but the final segment, where seeking further information was preferred (seek info = 74%,

don't extricate = 26%, $p<.001$, $g = .24$). Correspondingly, the majority of participants responded without suspicion in all segments of the scenario. Figures 14 and 15 display the patterns of responding.

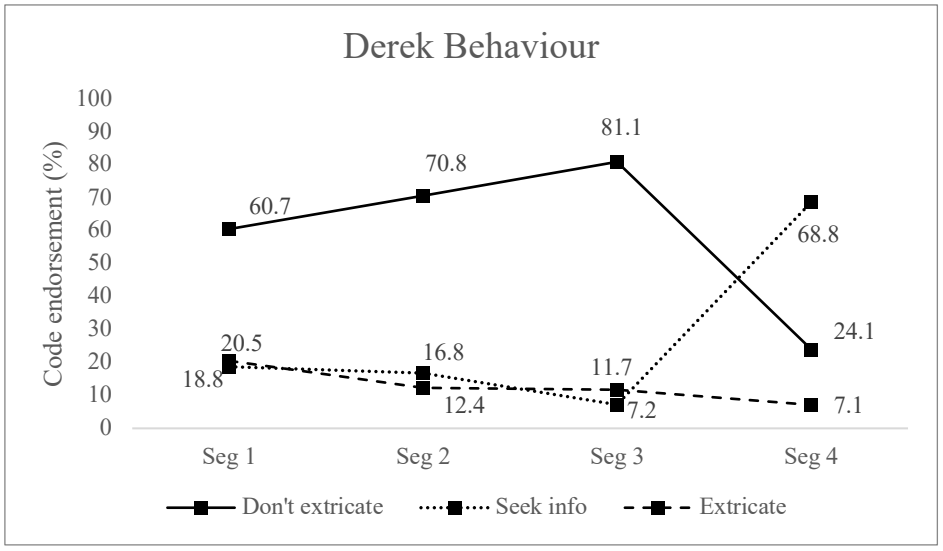


Figure 14. *Line graph representing the percentage of each behaviour code at each segment in the Derek scenario*.

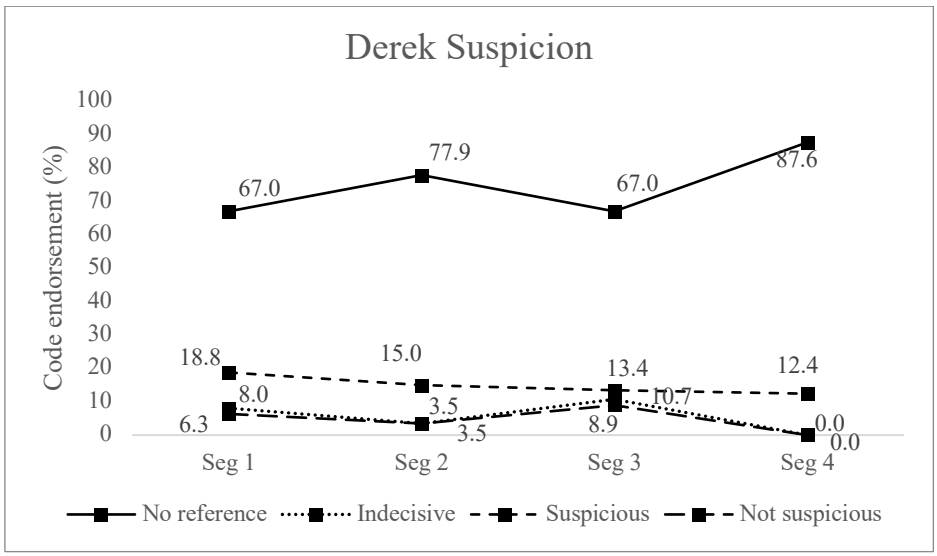


Figure 15. *Line graph representing the percentage of each suspicion code at each segment in the Derek scenario*.

Discussion

The paradigm development data indicated that, although the scenarios were broadly designed to become progressively more suspicious, variations in the subtlety of the cues at each step reduced the likelihood that suspicion would increase in a consistent, or linear, manner across all scenarios. Despite this, suspicions were raised by the majority of participants by the final segment of all crime scenarios, with the exception of the Andy scenario. The Aaron scenario, which insinuated that the protagonist stored child exploitation material, demonstrated the earliest recognition of suspicion by participants (61% at segment 3). This may reflect that the perceived gravity of the offence took precedence over potential alternative explanations, particularly as the dialogue did not explicitly state that the images were inappropriate. The Aaron scenario appeared to elicit the most emotive responses, with many participants reflecting on the abhorrence of the crime (e.g., “I would tell him he's a sick pervert and a liar. I'd punch him in the face then call the police”). By contrast, the Jia crime scenario, in which a young woman unknowingly applied for a fraudulent online university to maintain a student visa, elicited very little suspicion until the final segment when another character expressed concern with the situation. It is possible that the notion of a fraudulent online ‘diploma mill’ (Brown, 2006) is less recognisable than non-cybercrimes, which would be consistent with the high rates of international cybercrime victimisation. In the Andy scenario, the protagonist accepted a job as a driver for individuals who were depicted as being of a questionable nature (e.g., needing transport to a liquor store and courthouse); however, some cues required stereotypical judgements of character and may have been perceived as coincidental (e.g., the pick-up was in a run-down suburb and the client was wearing ripped clothing and a backpack). Indeed, some responses reflected this notion. For example, one participant responded, “Andy should pick her up. There is no dress code so how she dresses doesn't matter”. Additionally, the final segment in the Andy scenario involved a

young male being driven to a city building late at night and asking the protagonist to wait for a few minutes. At this point, many participants indicated that they could not in good conscience leave a child alone late at night. It is possible that, had the final antagonist been depicted as a fully grown adult, participants would have been less willing to oblige. Finally, in the Amy crime scenario, where a customer attempts to run a quick-change scam on two supermarket cashiers, many participants indicated that they would seek the assistance of their manager or review their transaction before complying with the man's request. These behaviours were often expressed in lieu of any suspicion around the man's intentions, indicating that general workplace procedures acted as a safeguard for many participants in this situation.

In summary, while many cues were perceived as suspicious enough to warrant concern, participants often appropriately reasoned that the ambiguity of the situation necessitated some hesitation in responding suspiciously (e.g., "he could be a scammer or con, but I also could have made a mistake"). Therefore, both extricating behaviours and seeking further information in the crime scenarios could be broadly conceptualised as adaptive behavioural responses, and non-extricating behaviours as maladaptive responses. Similarly, indecisively referencing suspicion and feeling suspicious could be considered adaptive and making no reference to suspicion or stating that the situation was not suspicious could be maladaptive.

The non-crime scenarios revealed an encouraging pattern that most participants were not primed to expect suspicious activity in every scenario. While the Casey scenario elicited a significant proportion of suspicious responding by the final segments, this was not entirely surprising given the protagonist was mistakenly accused of stealing their friend's bicycle following several coincidental cues (e.g., buying a new bicycle from a private seller around the same time as her friend's was stolen). A clearer timeline, in which the protagonist purchased the bike prior to learning of her friend's misfortune, may have assisted in allaying

some suspicions. It is clear that the subtlety of the contextual cues was paramount in impacting perceptions of suspicion and may be an avenue for future experimental manipulation.

The current data highlight considerable variation in individual response patterns across scenarios. This variation rules out the possibility that a 'most appropriate' response—with which all others could be compared and, in turn, judged as either correct or incorrect—could be identified for each segment. Given that some clear differences, with at least moderate effect sizes, were apparent in the proportions of responding, there are several segment endings by when it could be expected that most participants should be suspicious (i.e., segments 3 to 5 in the Aaron scenario and segment 5 in the Amy scenario) or demonstrate extricating behaviours (i.e., Aaron segment 4, and Amy segments 2, 3, and 5). However, identifying some overall (i.e., across scenarios) cumulative 'behaviour' and 'suspicion' score on the task that indicates normative responding does not appear to be viable. Rather, it may be more appropriate to consider group differences in the proportion of participants who report each behaviour (e.g., declining, seeking information, accepting offer etc.), as well as patterns over time, when examining the possible vulnerability of autistic individuals.

One obvious limitation of the development work conducted thus far is the lack of potentially relevant information about respondents that may have shaped responding. For example, given the difficulties autistic individuals face in gaining employment (Gal et al., 2015; Hendricks, 2010; Taylor et al., 2015), it is not surprising that some research has found an overrepresentation of autistic individuals turning to online crowdsourcing platforms such as MTurk for income (Chandler & Shapiro, 2016). In a recent study using MTurk particpants, 2.65% of 302 respondents self-reported a diagnosis of autism. It is possible, therefore, that a number of participants in the current sample may have scored highly on a measure of autistic

traits, or poorly on a measure of ToM, which in turn may be reflected in the responses obtained. This limitation is addressed in future studies.

Thus far, however, the SAP paradigm appeared to be a viable way to capture rich data regarding people's behaviours and their recognition of suspicion across several problematic situations. It adequately discriminated between appropriate and overly suspicious responding by demonstrating that participants were not primed to expect criminal activity in every scenario, and was able to capture a broad range of responses due to the variability within and between scenarios. Before embarking on an examination of autistic individuals' ability to recognise potentially problematic interactions, however, the paradigm was further evaluated in a second study investigating the relationship between the detection of suspicious activity and ToM ability.

CHAPTER 3

Study 2

The second study served several purposes. First, it augmented the previous data by examining whether patterns of responding remained consistent when using a larger sample. Enlarging the sample also facilitated further investigation of the optimal ways to score responses for each segment and the viability of an overall composite score. In addition, the second study allowed preliminary exploration of a relationship between the detection of suspicious activity and ToM, and other potentially relevant variables. Although Brewer et al. (2023) found no relationship between autistic traits and the discrimination of suspicious cues, a measure of autistic traits was included to investigate whether such traits were related to performance on the SAP. Furthermore, although I hypothesised a relationship between ToM and SAP performance would be detected given adequate variability on an adult measure of ToM, it was acknowledged that meaningful relationships may not emerge when using only non-autistic adults. Previous research has demonstrated that high proportions of non-autistic adults score at or near ceiling on the ToM measure to be used (Brewer et al., 2017; Brewer et al., 2022).

Method

Participants

One hundred and fifty participants were recruited from Amazon Mechanical Turk to complete all tasks in online format. A power analysis conducted with G*Power 3.1 (Faul et al., 2017) suggested that a minimum sample of 123 participants would provide sufficient power (0.80) to detect a medium effect (0.25) at an alpha level of .05 in a bivariate correlation. Worker requirements were unchanged from the previous study. Twelve participants were excluded due to uninterpretable responses, leaving a sample of 138 participants (68 males, 70 females; age range 21-71, M = 39.41, SD = 11.75).

Materials

Suspicious Activity Paradigm (SAP). The Suspicious Activity Paradigm (SAP) was presented in identical form to that of Phase 2 in the first study.

Adult Theory of Mind task (A-ToM-Q). Theory of Mind (ToM) was assessed using a modified version of the Adult Theory of Mind task (A-ToM; Brewer et al., 2017), which consists of 12 naturalistic video vignettes, some of which were adapted from the Strange Stories task (Happé, 1994). Six 'social' scenarios of everyday interactions capture understanding of various characters' social intent, while six 'physical' scenarios require no social inferences. Within the social scenarios, participants are required to recognise several facets of social interaction including sarcasm, white lie, misunderstanding, persuasion, and faux pas. The original A-ToM task was validated using a large sample of 163 autistic and 80 non-autistic participants with IQ > 85. Evidence of clear discrimination between groups on the social but not physical items, as well as convergent, divergent, and criterion-related validity has been reported in several studies (Brewer et al., 2017; Brewer et al., 2019; Young & Brewer, 2020). While the original task required open-ended responding, a forced-choice version (A-ToM-Q; Brewer et al., 2022) was used for the present research due to its short administration time and automated scoring. Like its predecessor, the A-ToM-Q has demonstrated discriminant, concurrent, and divergent validity (Brewer et al., 2022). For the current study, participants were only required to complete the social (i.e., the ToM) scenarios (which range in duration between 24 and 62 seconds) to minimise the substantial time demands for participants involved in completing all measures. After viewing each stimulus video, participants were presented with four, forced-choice response options and asked to respond as quickly and accurately as possible. Scores were classified as either incorrect or correct and ranged from 0 to 6. An example of one social scenario's script and the corresponding response options is shown in Table 10.

Table 10

Example stimulus script from the A-ToM-Q Social subscale and the corresponding scoring criteria

Social story (Bunnies)	
Script	Scoring
Two women sit in their living room discussing their bunnies: SUSIE: "So you know there is a lady coming over today to take a look at the rabbits." MRS SMITH: "That's good, because you know we can't keep them all." SUSIE: "I know." She looks sad as she picks up one of the bunnies and cuddles it. SUSIE: "I just love them so much. I can't bear the thought of anything bad happening to them. They're just so beautiful and cuddly." A girl approaches the house and knocks on the front door. The door opens to reveal woman 1 and woman 2: POTENTIAL BUYER: "Hi, I'm here to look at the bunnies." SUSIE: "Of course, come inside." Mrs Smith, Susie and the potential buyer are sitting in the living room. The potential buyer is cuddling one of the bunnies. POTENTIAL BUYER: "Oh they are all so cute. It's a shame they're all have males though, I was really looking for a female bunny." SUSIE: "Oh that is a shame. You know if I can't find a good home for them, I'm going to have to drown them."	Question to Participant: "Why does she say she will have to drown the rabbits? Please answer each of the following questions as quickly and accurately as you possibly can." **a) She is trying to make the person feel guilty so they will buy one of the rabbits. (correct)** b) She is trying to get the girl to buy one. (incorrect) c) She is unable to keep them all and if she can't she will have to kill them. (incorrect) d) She's a horrible person who hates rabbits. (incorrect)

Autism-Spectrum Quotient (AQ-50). The AQ-50 is a 50 item self-report screening tool that targets five areas considered to be related to the strengths and difficulties of individuals with autism spectrum disorder: social skill, attention switching, attention to detail, communication, and imagination (Baron-Cohen et al., 2001). The AQ-50 was designed as a

brief and straightforward measure for individuals with average IQ or above. It requires respondents to agree or disagree with a set of statements, providing an indication of where they fall on a spectrum of autistic characteristics. The AQ-50 has been widely used in autism research and was originally considered to have good discriminant validity, reasonable construct validity and good test-retest reliability (Baron-Cohen et al., 2001; Hoekstra et al., 2008). A score of 32 or above is considered a sufficient cut-off for identifying clinically significant levels of autistic traits, with 80% of autistic participants scoring above 32 compared with 2% of non-autistic participants (Baron-Cohen et al., 2001). Recent evaluations, however, have revealed evidence for improved factor structure among reduced-item versions (Lundqvist & Lindner, 2017). Nevertheless, for the present study, the AQ-50 was included over shortened versions to increase potential variability in scores as we did not specifically target an autistic sample.

Spot the Word – Second Edition (STW-2). The Spot-the-Word test (Baddeley et al., 1993) is a measure of verbal IQ that was designed to estimate premorbid intellectual ability. Participants are required to view 100 word pairs, one of which is a real word while the other is not (e.g., kitchen - harrick), and identify the real word. The task has been evaluated against the WAIS-IV intelligence scales so that a "proxy VCI" score can be derived from the STW total score (Baddeley & Crawford, 2012). STW-2 is time efficient, easily implemented online, and has demonstrated convergent validity with measures of fluid intelligence and vocabulary (Baddeley et al., 1993; Mackinnon & Christensen, 2007; Yuspeh & Vanderploeg, 2000).

Design and Procedure

This study was correlational in design, examining potential relationships between Theory of Mind, responses on the SAP, and autistic traits. Ethics approval was granted by the Human Research Ethics Committee. Participants were able to complete the study from a

location of their choosing but were required to participate using a computer. Any participants who attempted to complete the study on a mobile device were screened and could not continue with the survey. In the study instructions, participants were warned that participation was expected to take up to two hours, but that opportunity for short breaks between tasks was available. Approximate completion times were provided with the instructions of each task so that participants could decide whether to take a break prior to beginning the task. Participants were also warned not to close their browser when taking breaks. Participants were first asked to provide brief demographic information including age, gender, ethnicity, and profession and were required to complete an attention check. They then completed the SAP, AQ-50, the social scenarios of the A-ToM-Q, and the Spot the Word task.

Results

Inter-rater reliability

The process for calculating inter-rater reliability in Study 1 was repeated for Study 2. A research assistant who was unfamiliar with the project completed a training session before coding a small number of responses from Study 1. Responses and discrepancies were discussed before 20 participants' responses from the current study were coded independently and inter-rater reliability was calculated (see Appendix H, p. 191). Any discrepancies were resolved, and a second round of coding for 15 additional participants was conducted. The inter-rater reliability statistics for the second round of coding are displayed in Table 11.

The kappa statistic for some scenarios appeared low in comparison to the percentage of agreement between raters (e.g., the suspicion scale of the Derek scenario). A review of the data indicated that the imbalanced distribution of scores, where the vast majority of participants in the Derek scenario were not suspicious, may have influenced the kappa coefficient, a phenomenon known as the "kappa paradox" (Feinstein & Cicchetti, 1990). To circumvent the kappa paradox in this case, an alternative coefficient, Gwet's AC1 (Gwet,

2008; Wongpakaran et al., 2013), was calculated for the suspicion scale of the Derek scenario. Inter-rater reliability using Gwet's AC1 was .83.

Table 11.

Second round of inter-rater reliability for coding on each scenario

Scenario	Behaviour		Suspicion	
	% Agreement	Kappa	% Agreement	Kappa
Aaron	88	.86	93.3	.89
Andy	97	.94	89	.72
Amy	86.7	.83	84	.75
Jia	93.3	.90	93.3	.81
Casey	84	.80	85.3	.76
Derek	95	.91	85	.54

Comparison of SAP results from Study 1 to Study 2

To compare responses on the SAP from Study 1 to Study 2 and allow for Cochran-Q (Cochran, 1950) analyses which require a dichotomous outcome, the proportions of each response were plotted using the dichotomous codes 'adaptive' and 'maladaptive' for behavioural responses, and 'suspicious' or 'not suspicious' for suspicion recognition. The rationale for using the broad 'adaptive' and 'maladaptive' categories was that several behaviours fall under an umbrella of responses that would be considered adaptive in a potentially problematic situation, while others would be considered maladaptive. Behaviours in the adaptive category were taking action, denying/declining an offer or request, extricating, seeking further information, and involving authorities. Conversely, those considered maladaptive included accepting an offer/complying with a request, being conversational, and continuing as they were. Although not all segments in all scenarios follow this pattern (i.e., the 'maladaptive' behaviours would not be considered maladaptive in non-crime scenarios where there is no criminal activity), the combination of behaviours are referred to under these terms across all scenarios for consistency. Examination of the proportions of each code

revealed remarkably similar patterns of responding for each scenario across the two studies. To illustrate, Figures 16 and 17 display the patterns of responding for the Aaron and Amy scenarios in Study 1 and 2. The remaining comparison graphs can be found in Appendix I (pp. 192-194) and the Cochran-Q and Chi-Square analyses for Study 2 SAP responses in Appendix J (pp. 195-196).

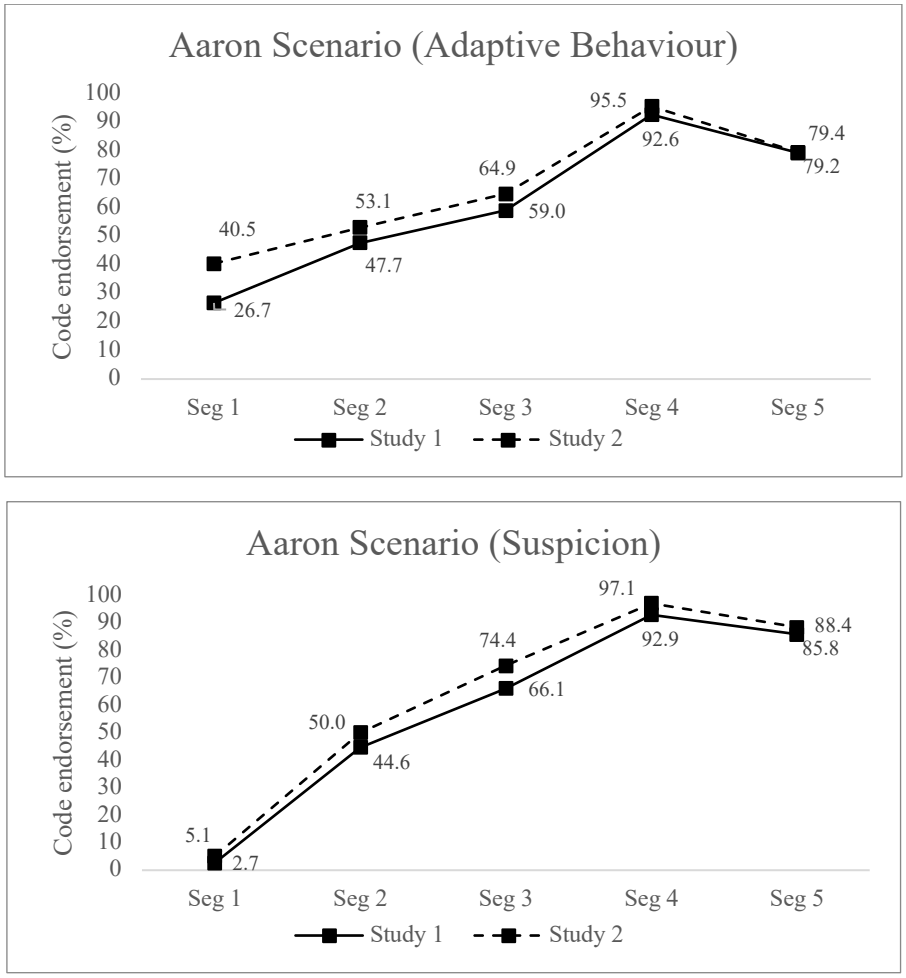


Figure 16. *Line graphs representing the percentage of adaptive behaviours and suspicion in the Aaron scenario in Study 1 and 2*.

Due to the similarity in patterns of responding on the SAP in Study 1 and 2, the data for the two studies have been combined for each scenario to provide a more stable estimate of what can be considered as an indication of normative performance on that task scenario. In future studies, these proportions will become the hypothesised proportions with which autistic and non-autistic groups can be compared.

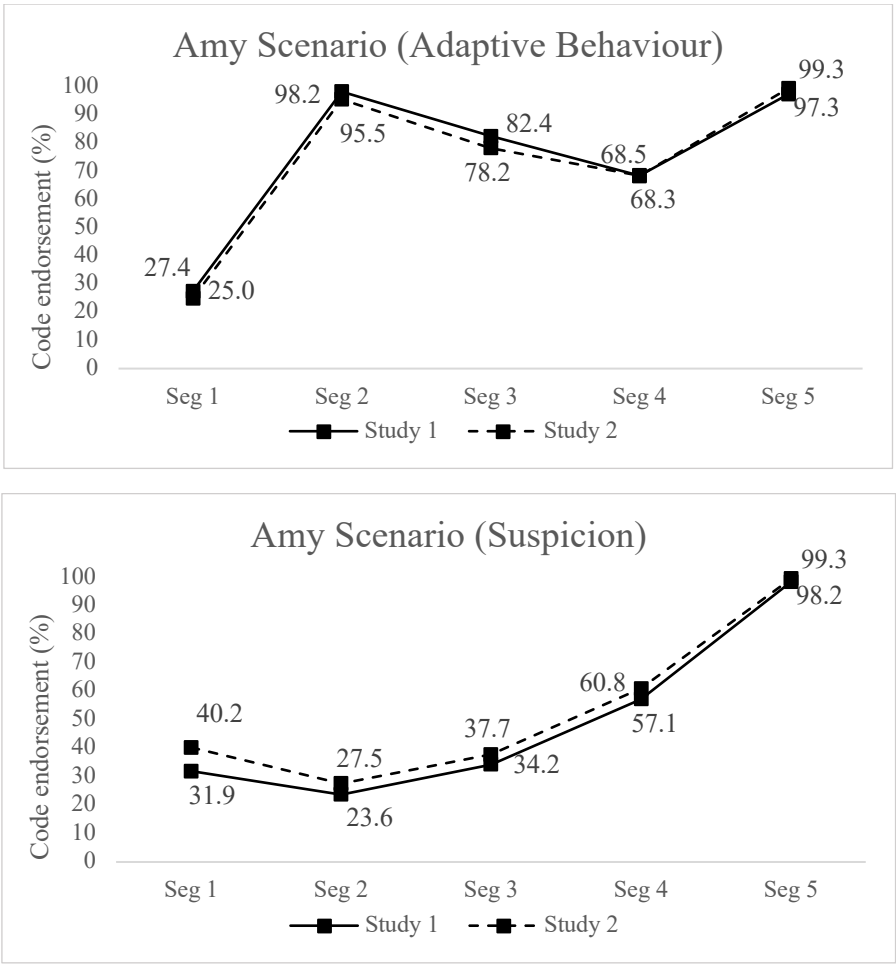


Figure 17. *Line graphs representing the percentage of adaptive behaviour and suspicion in the Amy scenario in Study 1 and 2.*

Normative SAP responses

Due to the similarity in responding in Study 1 and Study 2, the combined or normative patterns of responding deviate very little from those reported in Chapter 2. Nevertheless, given the different patterns across scenarios, it is informative to note the combined response patterns for each scenario if only to reinforce the SAP's sensitivity to variations in suspicion and individuals' behavioural responses within and across scenarios.

The following analyses were conducted with different sample sizes across segments due to excluded data. Responses that were excluded fell into the 'non-behavioural' or 'other' categories, or those that were missing due to audio failure. Excluded data made up 4.93% of all behavioural responses and 1.14% of all suspicion responses (total responses = 15,060).

Response frequencies and excluded data for each segment/scenario are reported in Appendix K (pp. 197-198).

Cochran's Q tests (Cochran, 1950) were used to determine whether the percentage of adaptive and maladaptive behaviours and suspicion, or absence thereof, differed across segments. As expected, given the fluctuations in responding highlighted in Figures 17 to 22, these were all significant. The Cochran-Q and post-hoc Dunn tests with Bonferroni correction appear in Appendix L (p. 199). Because the content of the scenarios did not always lead to increased suspicion at every segment, it was more informative to assess whether there were significant differences in the codes endorsed at each step individually. To do so, chi-square goodness-of-fit tests were run using the same dichotomous outcomes. Based on the absence of existing data, analyses were run assuming that the likelihood of each code being endorsed was 50%.

Crime scenario 1 (Aaron). Figure 18 displays the pattern of behavioural and suspicious responding for the Aaron scenario. Valid behavioural responses ranged between 236 and 242, while suspicion responses ranged between 245 and 251 (total N = 251). Chi-square analyses revealed a consistent preference for adaptive behaviours and suspicion from segment 3 onwards (see Table 12 for the relevant chi-square statistics).

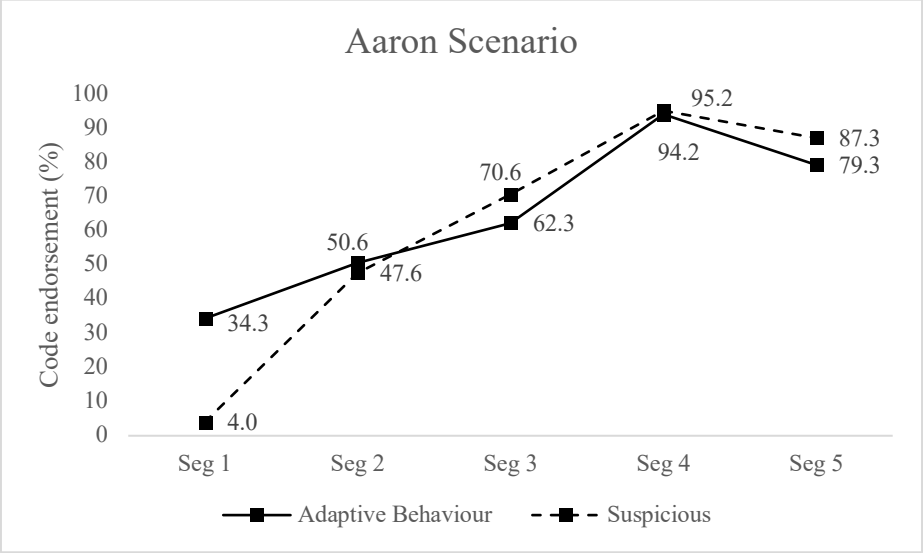


Figure 18. *Line graph representing the percentage of adaptive behaviours and suspicion in the Aaron scenario for studies 1 and 2 combined*.

Table 12.

Chi-square analyses for the Aaron scenario.

	N	One-Sample Chi Square	Cramer's *V*
Behaviour			
Segment 3	236	χ^2 (1) = 14.25, *p*<.001	.25
Segment 4	240	χ^2 (1) = 187.27, *p* <.001	.88
Segment 5	242	χ^2 (1) = 83.32, *p* <.001	.59
Suspicion			
Segment 3	245	χ^2 (1) = 41.64, *p* <.001	.41
Segment 4	250	χ^2 (1) = 204.30, *p* <.001	.90
Segment 5	251	χ^2 (1) = 139.32, *p* <.001	.75

Crime scenario 2 (Andy). Valid responses on the Andy scenario ranged from 236 to 246 (behaviour) and 246 to 251 (suspicion). Despite significant changes in the proportions of responses from segment to segment in both behaviour, $Q(5) = 128.58$, $p < .001$, and suspicion, $Q(5) = 165.84$, $p < .001$, there was no clear trend toward an increase in adaptive or suspicious responding (see Figure 19 for descriptive statistics). Suspicion was demonstrated by most participants in segments 5, χ^2 (1) = 12.05, $p < .001$, and 6, $\chi 2$ (1) = 13.46, p <.001, albeit by small margins (*V*=.22 and .23, respectively).

Crime scenario 3 (Amy). Valid responses on the Amy scenario ranged from 228 to 249 (behaviour) and 247 to 250 (suspicion). Adaptive behaviours were more likely to be endorsed from segment 2 onwards (see Figure 20 for descriptive statistics) with moderate to large effects: segment 2: χ^2 (1) = 213.05, p<.001, *V*=.93, segment 3: χ^2 (1) = 87.24, *p*<.001, *V*=.60, segment 4: χ^2 (1) = 30.95, $p < .001$, *V*=.37, and segment 5: χ^2 (1) = 227.26, $p < .001$, *V*=.97. As this scenario involved a financial scam in a supermarket, common adaptive behaviours in this scenario included calling over a manager or double checking the till for the amount given. Suspicion was aroused more slowly, with a significant majority of participants first becoming suspicious at segment 4 (χ^2 (1) = 8.20, $p < .001$, *V*=.18).

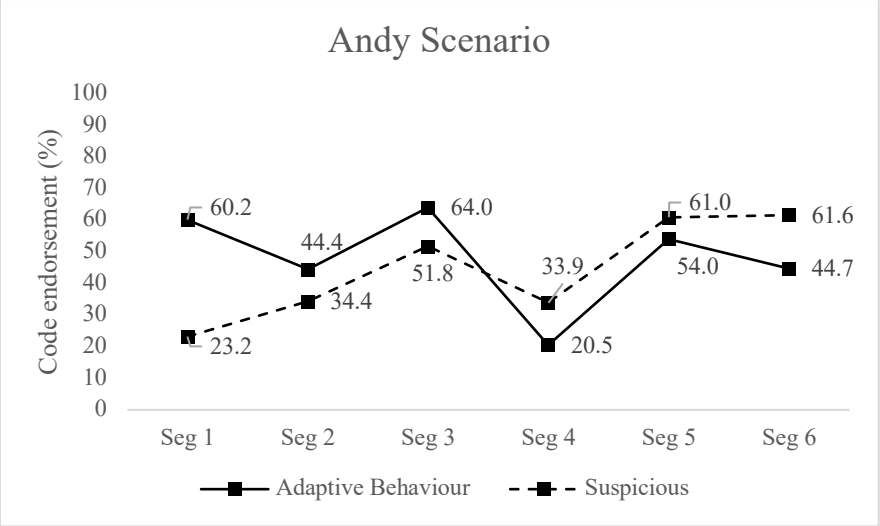


Figure 19. *Line graph representing the percentage of adaptive behaviours and suspicion in the Andy scenario for studies 1 and 2 combined.*

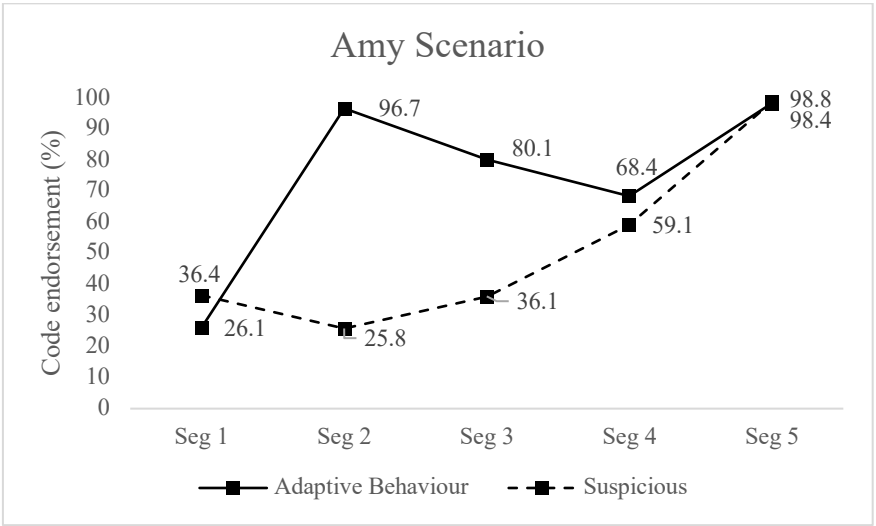


Figure 20. *Line graph representing the percentage of adaptive behaviours and suspicion in the Amy scenario for studies 1 and 2 combined.*

Crime scenario 4 (Jia). Valid responses in the Jia scenario ranged from 221 to 242 (behaviour) and 242 to 250 (suspicion). Although suspicion was generally only recognised by the final segment, χ^2 (1) = 237.15, p <.001, V = .45, adaptive behaviours were significantly favoured by segment 4, χ^2 (1) = 181.35, p <.001, V = .88. Adaptive behaviours at segment 4 included further attempts to contact enrolment officers or teaching staff, which were often

endorsed in lieu of suspicion about the legitimacy of the university. Figure 21 displays the descriptive statistics.

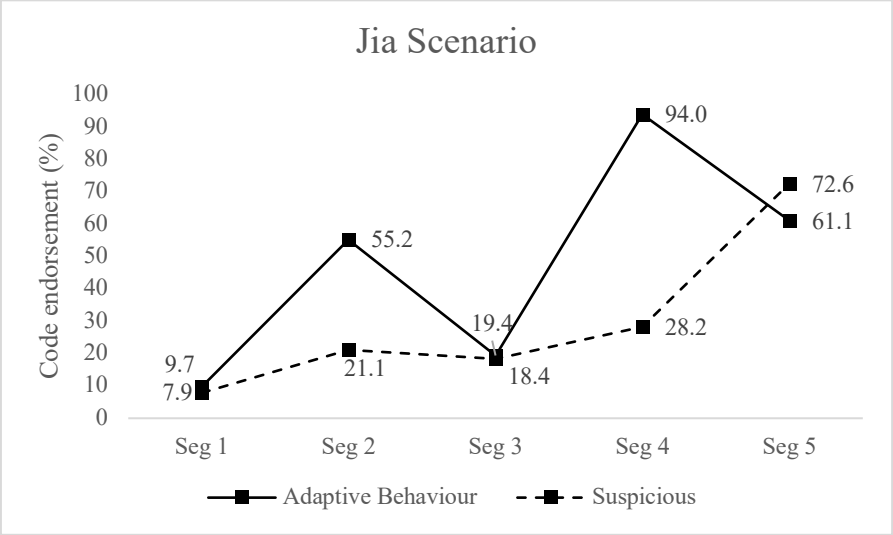


Figure 21. *Line graph representing the percentage of adaptive behaviour and suspicion in the Jia scenario for studies 1 and 2 combined.*

Non-crime scenario 1 (Casey). The Casey scenario produced the fewest valid responses due to a high occurrence of non-behavioural responding at segment 4 (n = 49). Responses in this category were often related to participants feeling a certain way (e.g., "I would feel worried I had purchased Ian's stolen bike") rather than how they would react behaviourally to the situation. Despite the absence of criminal activity in this scenario, suspicion and adaptive behaviours increased linearly, with both becoming the most common responses in segment 4, behaviour: χ^2 (1) = 28.88, p <.001, V=.38, suspicion: χ^2 (1) = 10.82, p <.001, V=.25, and segment 5, behaviour: χ^2 (1) = 167.23, p <.001, V=.83, suspicion: χ^2 (1) = 10.82, p <.001, V=.25. Figure 22 displays the descriptive statistics.

Non-crime scenario 2 (Derek). The Derek scenario produced the largest number of valid responses (behaviour = 245 to 250, suspicion = 249 to 251). It demonstrated that participants were not simply reporting suspicious behaviour because they were primed to do so by other scenarios, as the majority of participants did not endorse suspicion at any point

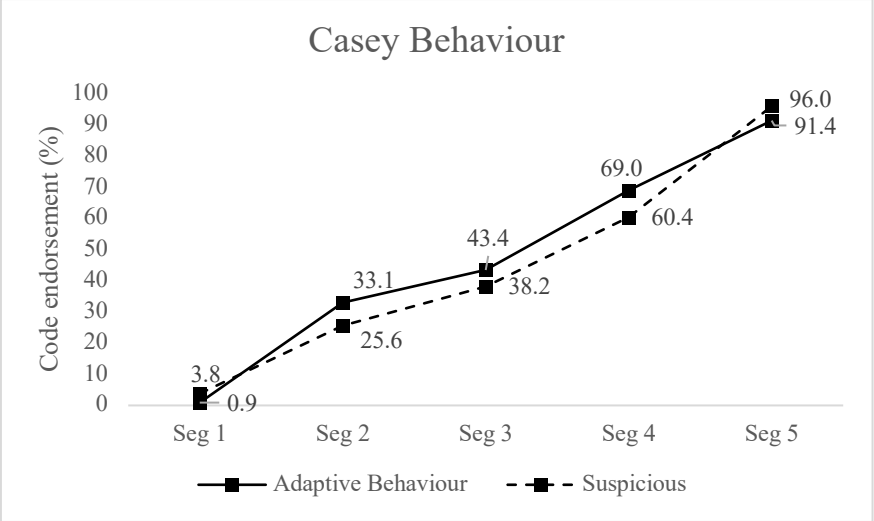


Figure 22. *Line graph representing the percentage of adaptive behaviour and suspicion in the Casey scenario for studies 1 and 2 combined.*

in the scenario (see Figure 23 for descriptive statistics). While 79.5% of participants reported an adaptive behaviour at segment 4, these included seeking information from friends and family about the wellbeing and whereabouts of the person in question.

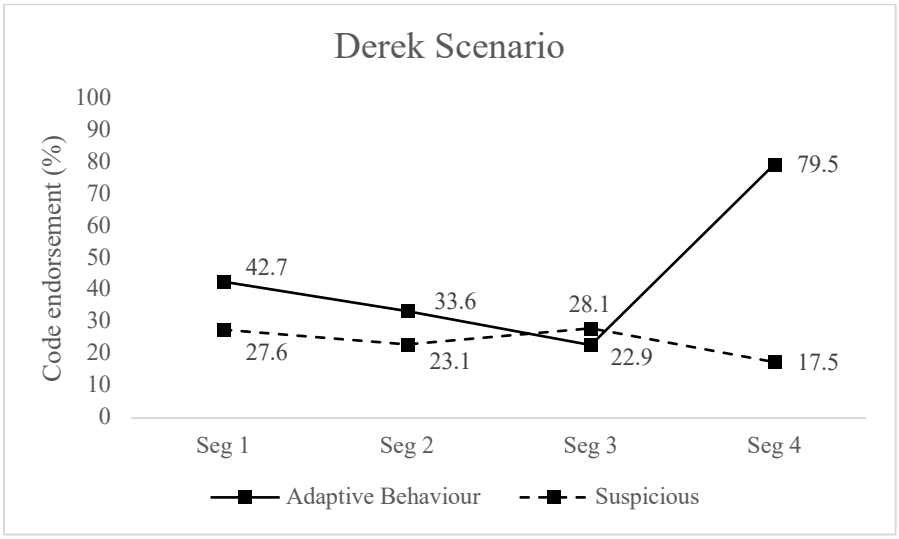


Figure 23. *Line graph representing the percentage of adaptive behaviour and suspicion in the Derek scenario for studies 1 and 2 combined.*

ToM and the Detection and Response to Suspicious Activity

As measures of ToM and IQ were not included in Study 1, the following analyses were run using the sample from Study 2 only (N=138). Consistent with the normative A-ToM and

A-ToM-Q data previously reported for non-autistic adults (Brewer et al., 2017, 2022), descriptive statistics revealed that many participants performed at ceiling on the social scenarios of the A-ToM-Q, with scores heavily negatively skewed (Mdn = 5). See Table 13 for descriptive statistics of all comparison measures. To examine the relationship between ToM and responses to the SAP, point-biserial correlations were run for each segment in each scenario—acknowledging that the severe range restriction on the A-ToM-Q for this sample mitigated against detecting meaningful relationships. The non-parametric coefficient, Spearman's Rho (Schober et al., 2018), is displayed due to data non-normality. Across all scenarios, there were no significant correlations between ToM and SAP responses, with two exceptions. The Andy scenario revealed a significant correlation between ToM and adaptive behaviour at segment 5, $\rho = .19$, p=.030, and between ToM and suspicion at segment 1, $\rho = .28$, p=.002, indicating that those with higher A-ToM-Q scores were more likely to endorse an adaptive behaviour or be suspicious at the respective points, although the associations were weak to moderate (Schober et al., 2018). Table 14 displays the point point-biserial correlations for behaviour and suspicion for each scenario.

Due to the impact of score distribution on point-biserial correlations (Brewer & Wells, 2006; Juslin et al., 1996) and the proportion of participants performing at ceiling on the A-ToM-Q, plots were created to display the ToM scores of people who endorsed adaptive behaviours or suspiciousness at each segment. This was done in two ways: (1) ToM for

Table 13.

Descriptive statistics for comparison measures in Study 2

	N	Min	Max	Mean	SD	95% CI
AQ50 Total	138	2	38	19.98	7.34	18.76, 21.20
AToM Social	138	2	6	5.14	.876	4.99, 5.29
STW Total	138	34	98	73.25	11.26	71.37, 75.13
Proxy VCI	138	75	121	103.22	8.09	101.87, 104.57

Table 14

Point biserial correlation coefficients (Spearman's Rho) for the SAP and A-ToM-Q

	Behaviour			Suspicion		
Segment	Rho	*p*	95% CI	Rho	*p*	95% CI
Aaron						
Segment 1	.05	.562	-.13, .23	.03	.729	-.14, .20
Segment 2	.12	.178	-.06, .29	.13	.149	-.05, .29
Segment 3	.09	.300	-.09, .26	.15	.096	-.03, .31
Segment 4	.02	.812	-.16, .20	-.07	.445	-.24, .11
Segment 5	.13	.120	-.04, .30	.10	.244	-.07, .27
Andy						
Segment 1	.12	.159	-.05, .29	.28	.001	.11, .43
Segment 2	.12	.173	-.06, .29	.20	.019	.03, .36
Segment 3	.04	.655	-.14, .21	.09	.323	-.09, .25
Segment 4	-.01	.942	-.18, .17	.14	.111	-.04, .30
Segment 5	.19	.030	.01, .35	.14	.110	-.04, .30
Segment 6	.05	.551	-.12, .22	-.06	.461	-.23, .11
Amy						
Segment 1	.09	.282	-.08, .26	.14	.103	-.03, .31
Segment 2	.01	.918	-.17, .18	.15	.087	-.03, .31
Segment 3	.09	.324	-.09, .26	.16	.063	-.01, .32
Segment 4	.03	.740	-.15, .21	.07	.430	-.11, .24
Segment 5	.03	.708	-.14, .21	.03	.716	-.14, .20
Jia						
Segment 1	-.01	.916	-.18, .17	.03	.710	-.14, .21
Segment 2	.03	.697	-.14, .21	.03	.753	-.15, .20
Segment 3	-.04	.658	-.21, .14	-.05	.590	-.22, .13
Segment 4	-.08	.376	-.25, .10	-.08	.351	-.25, .09
Segment 5	.17	.068	-.02, .34	.09	.302	-.09, .26
Casey						
Segment 1	-.04	.677	-.22, .15	-.02	.791	-.20, .16
Segment 2	-.04	.692	-.21, .14	.01	.922	-.16, .18
Segment 3	-.12	.171	-.29, .06	-.05	.530	-.22, .12
Segment 4	.01	.953	-.19, .20	-.02	.832	-.19, .15
Segment 5	.05	.594	-.13, .22	.10	.229	-.07, .27
Derek						
Segment 1	.08	.344	-.09, .25	.09	.298	-.08, .26
Segment 2	.02	.772	-.15, .20	-.02	.790	-.19, .15
Segment 3	-.09	.287	-.26, .08	-.06	.502	-.23, .12
Segment 4	.06	.478	-.11, .23	-.03	.704	-.20, .14

participants who reported suspicion or adaptive behaviours at any segment in the scenario, and (2) ToM for participants who responded suspiciously or adaptively for the first time only. It was anticipated that those with ToM difficulties would take longer to respond adaptively

and report suspicion. Again, the likelihood of discovering any differences was hindered by the limited variance in A-ToM-Q performance, with no clear negative trend in scores across the segments. Figure 24 contains the plots for the Jia scenario while all other scenarios can be found in Appendix M (pp. 200-202).

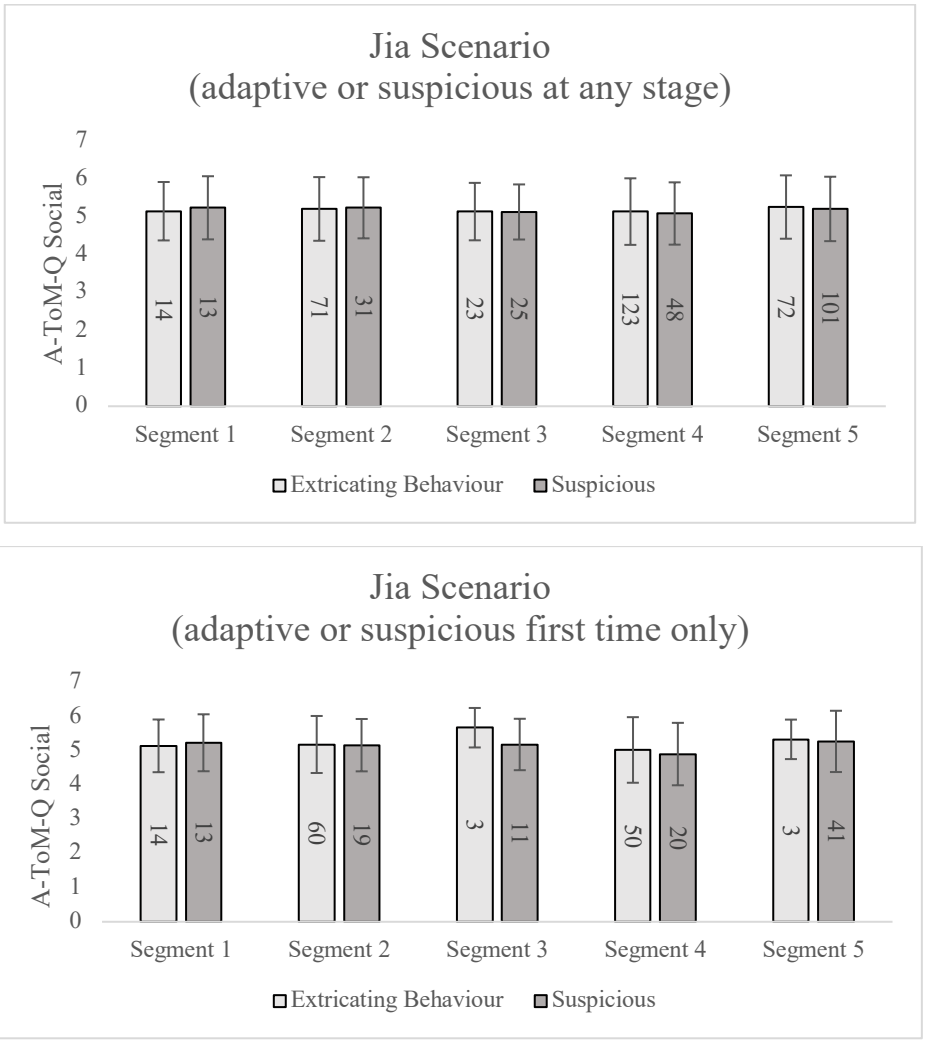


Figure 24. *Plots displaying the mean A-ToM-Q Social score (with standard deviation error bars and participant numbers) for participants who responded suspiciously or with an adaptive behaviour at each segment in the Jia scenario*.

Relationship between Autistic Traits and the Detection and Response to Suspicious Activity

The above analysis approach was repeated to investigate the relationship between autistic traits and SAP performance. There appeared to be no significant relationships

between autistic traits and behavioural or suspicious responding, with the exception of the Casey scenario, in which there were significant positive correlations between behaviour and autistic traits at segment 3, ρ = .24, *p*=.011, and suspicion and autistic traits at segments 3, ρ = .24, *p*=.005, and 4, ρ = .26, *p*=.002. This indicated that those with more autistic traits were more likely to respond adaptively or be suspicious at those points. All point-biserial correlation tables can be found in Appendix N (p. 203). Additionally, a correlation was run to determine whether those with lower A-ToM-Q scores demonstrated more autistic traits. This was nonsignificant.

The Relationship Between Intelligence and Detecting and Responding to Suspicious Activity

Further point-biserial correlations were run to investigate the relationship between proxy VCI scores, derived from the Spot the Word task, and SAP performance. Six participants were identified as having a proxy VCI score below 85, which was designated as the cut-off due to falling 1 standard deviation below the mean (Wechsler, 2008). Point biserial correlations were first run including all participants and then repeated with the six identified participants excluded to identify whether intelligence remained a significant predictor of SAP performance in their absence. The point biserial correlations revealed that in five out of six scenarios, at least one segment produced a significant relationship between proxy VCI and adaptive behaviour or suspicion. The non-crime scenario 'Casey' was the only scenario to produce no significant relationship with or without the excluded participants. Table 15 displays the point-biserial correlations with all participants included. When the six participants with a proxy VCI of below 85 were removed, the relationships between proxy VCI and SAP performance remained in several segments across the Aaron, Andy, Jia, and Derek scenarios. These are displayed in Appendix O (p. 204). Plots displaying the mean

proxy VCI score and standard deviation for participants who responded suspiciously or with an adaptive behaviour in each scenario are displayed in Appendix P (pp. 205-209).

Table 15

Point biserial correlation coefficients (Pearson) for the SAP and VCI with no exclusions

	Behaviour			Suspicion		
Segment	*r*	*p*	95% CI	*r*	*p*	95% CI
Aaron						
Segment 1	.18	.043	.01, .34	.03	.686	-.13, .20
Segment 2	.24	.007	.07, .40	.29	.001	.13, .44
Segment 3	.25	.004	.08, .40	.24	.006	.07, .39
Segment 4	.26	.003	.09, .41	.09	.296	-.08, .25
Segment 5	.19	.025	.02, .35	.19	.026	.02, .35
Andy						
Segment 1	.20	.018	.04, .36	.21	.016	.04, .36
Segment 2	.25	.003	.09, .41	.23	.008	.06, .38
Segment 3	.05	.572	-.12, .22	.23	.006	.07, .39
Segment 4	.00	.983	-.17, .17	.17	.047	.00, .33
Segment 5	.24	.007	.07, .39	.18	.031	.02, .34
Segment 6	.07	.429	-.10, .23	-.02	.795	-.19, .15
Amy						
Segment 1	-.05	.541	-.22, .12	-.03	.695	-.20, .13
Segment 2	.01	.923	-.16, .18	-.04	.652	-.20, .13
Segment 3	.12	.169	-.05, .28	-.08	.338	-.25, .09
Segment 4	.07	.426	-.10, .24	.05	.539	-.12, .22
Segment 5	-.03	.746	-.20, .14	.20	.017	.04, .36
Jia						
Segment 1	.07	.393	-.10, .24	.21	.013	.05, .37
Segment 2	.18	.038	.01, .34	.21	.017	.04, .36
Segment 3	.10	.231	-.07, .27	.23	.007	.07, .38
Segment 4	.07	.396	-.10, .24	.12	.171	-.05, .28
Segment 5	.05	.605	-.13, .23	.17	.049	.00, .33
Casey						
Segment 1	.13	.166	-.05, .30	.10	.284	-.08, .27
Segment 2	-.10	.249	-.27, .07	-.06	.489	-.23, .11
Segment 3	.02	.836	-.15, .19	.04	.650	-.13, .20
Segment 4	-.02	.836	-.21, .17	.08	.350	-.09, .24
Segment 5	.07	.424	-.10, .24	.10	.222	-.06, .27
Derek						
Segment 1	.31	<.001	.15, .46	.22	.009	.06, .38
Segment 2	.22	.010	.05, .37	.16	.060	-.01, .32
Segment 3	-.06	.522	-.22, .11	-.04	.620	-.21, .13
Segment 4	-.12	.157	-.28, .05	.04	.657	-.13, .20

Discussion

Study 2 extended Study 1 by investigating response patterns on the SAP in an online sample from the general population. The findings of Study 1 were broadly replicated with a larger sample, thus creating a snapshot of what might be considered pseudo-normative responses to the task. This also allowed for further investigation into whether a composite of scores across scenarios was a viable option, rather than individual reporting of scenarios. The replication of previous results did not support the viability of an overall composite score. In addition to further exploration of the SAP, Study 2 investigated potential relationships between SAP responding and measures of ToM, autistic traits and verbal IQ. Unsurprisingly, given the (anticipated) limited variability in performance on the ToM task that characterises non-autistic adults, these results did not provide evidence of a relationship between ToM and detection and response to suspicious behaviour. Future study directions to address this will be discussed in the Conclusions section.

Normative Performance on the SAP

As discussed in Chapter 2, response patterns on the SAP varied across the six scenarios. Therefore, it is useful to consider each scenario in isolation when evaluating SAP responses. The Aaron crime scenario (depicting a request for storage of child pornography) elicited the most consistent increase in suspicion and corresponding adaptive behaviours. Conversely, the Andy crime scenario (where the protagonist is set-up as a getaway driver) demonstrated non-linear patterns of responding, with an even distribution of adaptive and maladaptive behaviours in the final two segments. Despite the final two segments producing significantly more suspicious responses than non-suspicious responses, the differences were small. Nevertheless, while the Andy scenario initially appeared to be of limited utility due to the mixed response distribution, it may still prove to be informative when autistic and non-autistic groups are compared in Study 3. It was also the only scenario to demonstrate any

relationship with ToM, with participants with better developed ToM more likely to respond adaptively at the fifth segment, and more likely to report suspicion in the first segment. This is a promising finding considering the constraints of the ToM scores on the ability to reveal meaningful relationships. The Amy scenario (a cash scam in a supermarket) depicted a steady increase in suspicion, with most participants reporting suspicious behaviour by the fourth segment. Much like the first study, this was despite significant and large proportions of adaptive behaviours being endorsed as early as segment 2. This resulted from adaptive behaviours (e.g., calling a manager over and double checking the till) being attributed to workplace procedures regardless of concerns about malicious intent. Similarly, the fourth crime scenario (Jia), which depicted an illegitimate university course, demonstrated a significant majority of adaptive behaviours at segment 4 that were incongruent with suspicion recognition. This was due to most participants indicating that they would continue seeking further information from university staff or attempt to visit the university in person, despite making no reference to the possibility of the university being fraudulent.

The non-crime scenarios produced differing response patterns. The Derek scenario, in which the protagonist was asked to store a friend's belongings while she is away on holiday, produced a maximum of 28.1% suspicious responses across each segment. This indicated that participants were not continually responding with suspicion because the crime scenarios had primed them to do so. The Casey scenario, conversely, demonstrated the most consistent increase in suspicion and adaptive behaviours of all scenarios. While the cues were generally considered coincidental (e.g., a friend's bike being stolen around the same time Casey had purchased a new one), the conclusion of the scenario, in which Casey is wrongfully accused of stealing the bike, was undoubtedly too suspicious. Because the response pattern for the Casey scenario represented what was expected from the crime scenarios, two options were considered regarding its use in future studies: that it be converted to a crime scenario by

alteration of the final segment, or that it be removed from the task completely. It was ultimately decided that the scenario should be removed to reduce the task completion time and associated costs, with minimal loss of informative content. Although it is possible that the removal of a non-crime scenario would increase the chance of suspicious responding due to fewer instances of disconfirming evidence (i.e., the final segments that confirm no crime has occurred), this should be reflected in responses on the Derek scenario.

Relationship between ToM and SAP Performance

Predictions about the relationship between ToM and detection and response to suspicious behaviour were not supported by the current data. This was unsurprising given the limited variability in performance on the ToM task. No meaningful differences were revealed, except for the Andy scenario in segments 1 and 5. Notably, the first segment of each scenario was not designed to be suspicious. In the case of Andy, it involved Andy receiving a job offer from a long-term acquaintance after disclosing his difficulty finding work. The job offer involved driving others around with limited additional information. Perhaps participants with better-developed ToM were more cynical of this offer as they were able to foresee potential dangers associated with the job. This finding, in the absence of significant relationships in other scenarios, may lend credence to the importance of ToM when cues are subtle. Importantly, A-ToM-Q performance in the current study (M = 5.14, SD = .88) was similar to that of the non-autistic sample in the original Brewer et al. (2022) study (M = 5.27, SD = .81), and studies that have since used the task (Brewer et al. 2023). In those studies, the comparison of non-autistic and autistic samples revealed consistent significant group differences in performance on the A-ToM-Q task, with the autistic samples characterised by greater ToM difficulties. Thus, the inclusion of a comparison group of autistic individuals would be expected to increase overall ToM variability, providing an opportunity to detect meaningful relationships between ToM and SAP performance.

However, given Brewer et al.'s (2023) findings of no group differences in 'dodginess detection' between autistic and non-autistic participants when using a signal detection theory approach, it is unknown whether the addition of an autistic sample will translate to any group differences on the SAP.

Autistic Traits and SAP Performance

The current data did not support the hypothesis that increased autistic traits would be associated with poorer performance on the SAP. There were no significant relationships between autistic traits and SAP performance, with the exception of the Casey scenario in which higher autistic traits were associated with adaptive behaviours at segment 3, and suspicion recognition at segments 3 and 4. In segment 3, Casey travelled to buy a second-hand bike after learning that her friend's bike was recently stolen. At that point, the connection between the two events was intended to be coincidental. In segment 4, her riding peers congratulated her on the new purchase and commented that the bike looked familiar. Perhaps those with increased autistic traits sought to find predictable patterns within the study (Baron-Cohen, 2008) and were more heavily primed to expect suspicious activity by the crime scenarios; however, this would likely also be reflected in the Derek scenario. Brewer et al. (2023) also found no significant association between autistic traits and detection of dodginess. While they used a 12-item version of the AQ-50, an evaluation of that measure demonstrated no loss in power from the reduced item version (Lundqvist & Lindner, 2017). Therefore, in the interest of reducing testing time, the AQ-12 will be used in the next study.

Verbal Intelligence and SAP Performance

Verbal intelligence was related to SAP performance in at least one segment in five of the six scenarios. Furthermore, when six participants were removed for demonstrating a proxy VCI below 85, the relationship remained in several segments and scenarios. This finding is in line with research that has linked performance on ToM tasks with intelligence

and verbal ability (Baker et al., 2014; Happé, 1994; Nilsson & de López, 2016). Despite this, group differences in ToM have remained after controlling for VCI in several studies (Brewer et al., 2017; Brewer et al., 2022). Considering the exclusive use of audio-recorded vignettes in the current research, it was expected that verbal ability would be associated with SAP performance.

Limitations

Despite targeting participants from several western countries, the current study was limited to an entirely US based sample due to the high proportion of Amazon Mturk workers residing in the US. A small number of participants indicated some difficulties understanding the Australian accents in the audio recordings. In future studies, the online crowdsourcing platform Prolific will be used in place of Amazon MTurk. Prolific provides access to a larger proportion of participants based in the UK and Australia. It also contains a subgroup of participants who have indicated a diagnosis of ASD in their demographic information, which will allow for group comparisons with acceptable sample sizes.

Furthermore, the ecological validity of the tool remains limited to some degree by the use of audio scenarios. Given the complexities of social interaction, there are undoubtedly many visual and contextual cues that cannot be conveyed through audio vignettes and yet individuals consider when they appraise a situation. It also collects rich data regarding people's anticipated behaviour and the reasoning behind their decision within the confines of the limited context provided. It is clear, however, that context and detail play a significant role in people's decision-making. For example, whether a character is described as an acquaintance or a close friend may impact their appraisal of the situation.

Conclusion and Future Directions

The present study was designed to further assess whether the SAP provides useful information regarding people's responses to suspicious behaviour, to provide a pseudo-

normative reference point for future studies, and to explore whether SAP responding might correlate with ToM ability in a sample of adults from the general population. Although these data did not provide evidence of an association between ToM and SAP performance, this was unsurprising given the limited variability in ToM demonstrated by the current sample.

Several limitations were identified that will be addressed in future studies. Primarily, the inclusion of an autistic sample will indicate whether the absence of autistic and non-autistic group differences in the discrimination of dodginess reported by Brewer et al. (2023) is reliable. It will also allow for larger variation in ToM ability which will increase the possibility of revealing meaningful group differences in SAP performance, or an overall relationship between ToM and the ability to detect and respond to suspicious behaviour. Previous studies using the A-ToM-Q have revealed significant group differences between non-autistic and autistic participants, with the non-autistic group performing at a level comparable to the current study (Brewer et al., 2022). It is therefore expected that the inclusion of an autistic sample will potentially increase the likelihood of lower scores, increasing the overall variability on the ToM measure.

CHAPTER 4

Study 3

The main objectives of this study were to examine whether autistic and non-autistic samples differed in the detection of, and response to, suspicious behaviour, and to examine further the relationship between ToM and the detection of suspicious behaviour. To address both aims, samples of autistic and non-autistic adults were recruited. This approach enabled (a) a comparison of the ability of autistic and non-autistic samples to detect and respond to suspicious behaviour, and (b) by introducing greater variability on the ToM measure, an enhanced opportunity to detect any meaningful relationship between ToM and SAP performance. In Chapters 2 and 3, I demonstrated the viability of the SAP as a method for capturing rich information regarding people's decision-making across a range of scenarios; however, it was also apparent that ceiling effects on the ToM measure limited the likelihood of finding meaningful associations between SAP performance and ToM. This was likely to be particularly relevant in those segments of crime scenarios in which the proportion of responses in one response category was very high (e.g., 95% adaptive or vice versa).

As previous research has shown consistent group differences between autistic and non-autistic adults and intra-group variability on the A-ToM (Brewer et al., 2017; Brewer et al., 2022), the measure of ToM used in the current study, it was anticipated that the inclusion of autistic and non-autistic samples would provide scope for potential relationships to emerge. In Chapter 3, weak relationships emerged in only one of the four crime scenarios. In that scenario (Andy), SAP response patterns indicated that the situation was not obviously suspicious, even by the final segment. The necessity to detect what clearly were quite subtle cues in this scenario perhaps meant that this scenario provided a more sensitive index of the influence of ToM despite the limited ToM variability.

Given the findings of previous psychometric evaluations of the A-ToM and A-ToM-Q tests (Brewer et al., 2017; Brewer et al., 2022), I hypothesised that, at the group level, autistic

participants would perform more poorly than non-autistic participants on this measure of ToM. Whether autistic adults would be less likely to detect suspicious behaviour and respond adaptively is difficult to predict. On the one hand, as already noted, it has often been suggested that ToM difficulties may render autistic individuals vulnerable to becoming involved in crime or being victimised. On the other hand, Brewer et al.'s (2023) signal detection study showed that, although discrimination of cues to suspicious or dodgy behaviour was associated with ToM, autistic and non-autistic adults did not differ on any indices of discriminability.

Third, given the combination of autistic and non-autistic samples should result in increased variability on the ToM measure, I hypothesised that that there would be a significant relationship between ToM and SAP performance, such that there would be positive correlations between ToM and the number of adaptive and suspicious responses.

The study also provided an opportunity to gather data on the relationships between the SAP, as a measure of the ability to detect suspicious behaviour, and potential independent indicators of convergent and divergent validity. Constructs that might be expected to relate to the ability to detect suspicious behaviour include deception detection (Williams et al., 2018), social intelligence (Silvera et al., 2001), social vulnerability (Pinsker et al., 2011), gullibility (Teunisse et al., 2020) and trust (Rotter, 1967). Given the importance of using quick, easily administered measures to ensure that the current battery of tasks was manageable for participants in what was already quite a demanding testing session, the Gullibility Scale (Teunisse et al., 2020) and Social Vulnerability Scale (Pinsker et al., 2011) were selected as candidate concurrent validity markers due to the limited number of items, self-report administration, and availability of existing validation data, all of which are discussed further in the Method section. As Teunisse et al. (2020) found that the 12-item Gullibility Scale was strongly associated with the Social Vulnerability Scale, but not the Interpersonal Trust Scale

(Rotter, 1967), the Interpersonal Trust Scale was also included as a measure of divergent validity. Although it might be assumed that trust and gullibility are conceptually related, research has shown that they are distinct constructs: trust is a general disposition toward believing the word of others whereas gullibility can only present itself in the face of untrustworthy cues. That is, highly trusting individuals may not be gullible because they are no less likely to recognise suspicious cues as they present. This notion has been supported by several studies (Rotter, 1980; Teunisse et al., 2020). In sum, I also anticipated that there would be negative associations between SAP responding and both gullibility and social vulnerability, whereas I did not expect a relationship between SAP responding and interpersonal trust.

Method

Participants

To detect a small to medium-sized difference in ToM between adult groups ($\alpha = .05$, power = .80, $df = 1$) as previously reported by Brewer et al. (2017) and Brewer et al. (2022), I targeted a sample of 198 participants (99 per group) (Faul et al., 2007). Two hundred and six participants were recruited from the online crowd-sourcing platform Prolific. The benefits of using Prolific over Amazon's Mechanical Turk include improved data quality (Peer et al., 2022) and the ability to target specific sample demographics. Including specific sample demographics allowed me to target participants who disclosed a diagnosis of autism spectrum disorder in their demographic information on the platform. Prolific also features a more significant proportion of participants from countries outside the United States, which allowed for broader representation; however, participation was still restricted to participants who indicated English was their first language. Two participants were excluded from the study due to inappropriate responses on the SAP (e.g., responding "don't know" to most segments or responses entirely out of context).

To allow a larger pool of autistic participants to access the study, participants who had previously participated in a study involving the A-ToM-Q were allowed to participate. To ensure potential practice effects did not bias the A-ToM-Q results, all participants' Prolific IDs were cross-checked with participant IDs from previous studies involving the A-ToM-Q. Fifty-three participants had pre-existing scores on the A-ToM-Q social scale, which we imputed for use in the current study. Five participants were excluded from analyses involving the A-ToM-Q due to technical difficulties that invalidated their scores.

The final autistic sample included 104 adults (46 male, 45 female, 11 non-binary, two preferred not to specify gender). Participants were asked to provide detailed information about their diagnosis, including confirmation of a formal diagnosis as a child or adult, the type of qualified professional who provided the diagnosis (e.g., general practitioner, psychiatrist, psychologist), and age at diagnosis. Age at diagnosis ranged from 2 to 59 years (M = 22.92, SD = 12.20, Mdn = 24.0) and diagnoses were most frequently provided by a psychiatrist (N = 53), psychologist (N = 24), or general practitioner (N = 11). Other diagnosticians included autism specialists (N = 6), paediatricians (N = 3), multi-disciplinary teams (N = 2), neurologists (N = 1), and social workers (N = 1). Three participants provided the name of their diagnosis (e.g., Kanner's Syndrome) or practitioner rather than the type of professional. The non-autistic sample comprised 100 participants (54 male, 45 female, and one non-binary). Table 16 presents the descriptive and inferential statistics for both groups on age and other comparison measures. Compared with the non-autistic group, the autistic group was, a little older, scored higher on the measure of autistic characteristics (AQ-12, Lundqvist & Linder, 2017) and, consistent with previous studies that used the A-ToM-Q (Brewer et al., 2023; Brewer et al., 2022), scored significantly lower on the ToM measure. Note, however, that a substantial proportion of each sample—as indicated by the median scores—scored at or near the maximum possible score of 6.

Table 16.
Descriptive and inferential statistics for comparison measures in Study 3

	N	Min	Max	Mean	SD	Median	95% CI
Autistic							
Age	104	18	61	32.70	9.41	31.0	30.87, 34.53
A-ToM-Q Social	101	0	6	4.40	1.52	5.0	4.10, 4.70
AQ12	104	0	12	9.11	3.28	10.0	8.47, 9.74
Proxy VCI	104	79	122	104.42	8.21	104.0	102.83, 106.02
Non-autistic							
Age	100	18	70	35.86	12.95	33.0	33.29, 38.43
A-ToM-Q Social	98	2	6	5.05	0.94	5.0	4.86, 5.24
AQ12	100	0	12	5.44	3.67	5.0	4.71, 6.17
Proxy VCI	100	73	121	101.06	10.07	102.0	99.06, 103.06

	t	*df*	*p*	*d*	95% CI	*U*	*p*
Age	1.99	180.40	.049	0.28	0.04, 0.56	4656.0	.197
A-ToM-Q Social	3.68	167.17	<.001	0.52	0.24, 0.80	3826.5	.004
AQ12	-7.52	202	<.001	1.05	0.76, 1.35	8027.5	<.001
Proxy VCI	-2.62	202	.010	0.37	0.09, 0.64	6212.5	.016

Verbal ability was estimated using the Proxy Verbal Comprehension Index score derived from the Spot the Word test (Baddeley & Crawford, 2012). To reduce the likelihood of participants not understanding the scenarios, participants who returned a proxy VCI score of less than 85 (Wechsler, 2008), were excluded. Seven participants (two autistic, five non-autistic) scored below 85 and were not included in further analyses. The autistic participants (M = 104.88, SD = 7.59, Mdn = 104.0, range = 89-122) demonstrated marginally higher verbal ability than the non-autistic participants (M = 102.18, SD = 8.98, Mdn = 102.0, range = 85-121), t (195) = -2.29, p = .023, d = 0.33, 95% CI [0.04, 0.61], U = 5697.50, p =.033.

Materials

Suspicious Activity Paradigm (SAP). Minor changes were made to the SAP. First, the Casey non-crime scenario was removed because its response patterns mirrored those in the crime scenarios, indicating inadequate discriminability. Second, due to the Casey scenario removal, the presentation order was changed. Previously, scenarios were presented in random order. There was no evidence to suggest a general increase in suspicion over time as the proportion of participants reporting adaptive behaviours and suspicion remained similar from

the first to last scenario displayed (i.e., 40% of responses were recorded as suspicious in the first scenario, and 45% were suspicious in the final scenario). To reduce the likelihood of this changing after removing the Casey scenario, the remaining non-crime scenario (Derek) was fixed as the third and central scenario. All crime scenarios were randomised around the non-crime scenario.

Adult Theory of Mind task (A-ToM-Q). The A-ToM-Q task (Brewer et al., 2022) described in Chapter 3 was presented in an identical format.

Spot the Word – Second Edition (STW-2). The STW-2 task (Baddeley & Crawford, 2012) described in Chapter 3 was also presented in an identical format.

Autism-Spectrum Quotient (AQ-12). The AQ-12 (Lundqvist & Lindner, 2017), a reduced item version of the AQ-50 used in the previous study, was used to reduce the study duration. The AQ-12 has been found to provide comparable explanatory power to the AQ-50 (Lundqvist & Lindner, 2017).

Gullibility Scale. The Gullibility Scale (Teunisse et al., 2020) is a recently developed 12-item, self-report tool that measures gullibility, defined by the authors as "an individual's propensity to accept a false premise in the presence of untrustworthiness cues" (p. 409). The scale contains two factors: persuadability and insensitivity to untrustworthy cues. Items include statements such as "I'm usually quick to notice when someone is trying to cheat me" and "My friends think I'm easily fooled", which are responded to on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). Scores range from 12 to 84. In their original validation studies (Teunisse et al., 2020), the final 12-item factor structure demonstrated good internal consistency ($\alpha = .92$) and was validated against several measures of convergent and discriminant validity. The authors reported a negative correlation between the gullibility scale and social intelligence (Grieve & Mahar, 2013), positive correlations between the gullibility scale and social vulnerability (Pinsker et al., 2011) and paranormal beliefs (Tobacyk, 2004),

but no relationship between gullibility and either interpersonal trust (Rotter, 1967) or Machiavellianism (Dahling et al., 2009). Test-retest reliability was assessed as stable (ICC = .80), and criterion validity was assessed by comparing a critical-thinking interest group (sceptics) with self-reported scam victims. Predictive validity was reported in a further study involving responses to simulated phishing email scams (George et al., 2020).

Social Vulnerability Scale-15 item version. The Social Vulnerability Scale (SVS-15; Pinsker et al., 2011) was initially developed for use with older adults with dementia or cognitive impairment to assess the risk of exploitation and vulnerability to financial abuse. It is an informant-based questionnaire that is typically completed by a partner or family member. The scale includes two primary factors (gullibility and credulity), and respondents rate the frequency with which their partner/family member engages in each item on a scale from 0 (Never) to 4 (Always). Scores range from 0 – 60. Example items include "Been persuaded to donate excessive sums of money to charities or other funds" and "Believes rumours that come from a questionable source". Internal consistency of the measure was good ($\alpha = .90$), and construct validity was demonstrated by a clinical sample (e.g., participants with vascular dementia, Alzheimer's, stroke, memory problems, or other neurological conditions) scoring significantly higher than a non-clinical sample. To my knowledge, the SVS-15 has not been evaluated using an autistic sample. For the current study, items were adapted for use with self-report (e.g., I believe rumours that come from a questionable source).

Interpersonal Trust Scale. The Interpersonal Trust Scale (Rotter, 1967) is a measure of dispositional trust that Rotter defined as "an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon" (p. 651). The scale contains 25 self-report items on a Likert scale from 1 (strongly agree) to 5 (strongly disagree). Scores range from 25 to 125. Items target trust

toward various social "objects", including parents, teachers, politicians, physicians, classmates, friends, and society. Internal consistency, test-retest reliability (r =.56 – .68), and relatively good construct and discriminant validity were demonstrated using a sociometric analysis where members of a university fraternity and sorority who lived together rated themselves and each other.

Procedure

Ethics approval was granted by the Human Research Ethics Committee. Participants were recruited via Prolific and completed the study on their computer. All participants were asked to confirm or deny a previous diagnosis of autism spectrum disorder and were screened from participating if their response did not correspond with their pre-existing demographic information on Prolific. Autistic participants then responded to additional questions regarding their diagnosis (i.e., age at diagnosis and type of qualified professional who provided the diagnosis) before all participants received the instructions and demographic questions used in Study 2. Approximate task completion times were updated for additional and modified tasks. The tasks were presented in fixed order due to the likely cognitive load of specific tasks, with more manageable tasks (e.g., short self-report questionnaires) breaking up the more complex and demanding tasks (i.e., SAP, A-ToM-Q, STW). Tasks were presented in the order of SAP, AQ-12, Gullibility Scale, A-ToM-Q, SVS-15, STW-2, and the Interpersonal Trust scale.

Results

Inter-rater Reliability

Inter-rater reliability for the current study was assessed using a single sample of responses coded by the author and a previously trained research assistant who had completed inter-rater coding for the analysis in Chapter 3. SAP responses from 50 participants (i.e., 250 total responses) were coded independently, and any discrepancies were discussed and resolved. Inter-rater reliability was acceptable for all scenarios (see Table 17 for reliability statistics). To illustrate the significant variability in individual responses captured by the

SAP, a selection of examples from various scenarios at each segment, and the subsequent codes, are included in Table 18.

Table 17

Inter-rater reliability for coding on each scenario (N=50)

Scenario	Behaviour		Suspicion	
	% Agreement	Kappa	% Agreement	Kappa
Aaron	86.8	.84	92.4	.87
Andy	89	.83	85	.76
Amy	87.2	.84	84	.75
Jia	90	.86	91.6	.86
Derek	90	.83	91	.83

Comparison of SAP Responses with Study 1 and 2 Data

SAP responses from the current study (N = 204) were compared with the 'pseudo-normative' response patterns derived from Studies 1 and 2 (N = 251) to confirm that the patterns were relatively stable. All participants were included in these comparisons as the previous study did not exclude participants based on proxy VCI scores. Consistent with the previous study, codes were combined into two dichotomous categories: adaptive behaviour and maladaptive behaviour, and suspicious or not suspicious. Participants who provided a response that fell into the 'non-behavioural' or 'other' categories, or experienced technical difficulties in a particular segment, were excluded from analyses. These responses made up 4.22% of the behavioural data and 1.65% of the suspicion data.

Perhaps unsurprisingly, given that the Casey non-crime scenario was not used in this study, descriptive statistics and chi-square analyses revealed a general increase in suspicious responding and adaptive behaviours when compared with the pseudo-normative data. However, the effect sizes for any significant differences were small[4] (Wiedmaier, 2017)

[4] Interpretation of the φ effect size is noted by Wiedmaier (2017) as .10 to .29 for a small effect, .30 to .49 for a medium effect, and .50 or greater for a large effect.

Table 18

Example responses from all scenarios at each segment of the SAP

	Behavioural response	Suspicion response
Segment 1		
Aaron	"Accept the situation and thank the manager." *(conversational/socially acceptable)*	"If everybody's hours are being cut then it makes sense that mine (Aaron's) is also cut. Perfectly understandable." *(no reference to suspicion)*
Andy	"I would ask for further details about the job and inquire thoroughly about all aspects." *(seek info/clarification)*	"His vague description of the job means that he is likely omitting something unsatisfactory." *(suspicious)*
Amy	"I would answer the man's question and ask if he needs anything else." *(conversational/socially acceptable)*	"My job is to chat to the customers." *(no reference to suspicion)*
Segment 2		
Jia	"I would enrol in the business class." *(go ahead/accept offer/comply with request)*	"It could help me find a job in the future, plus time in the classroom could be limited." *(no reference to suspicion)*
Derek	"Agree to hold them." *(go ahead/accept offer/comply with request)*	"Because I want to be helpful." *(no reference to suspicion)*
Andy	"I would politely decline, saying I had changed my mind about this job." *(decide against/decline offer/deny request)*	"I fear Carl deals drugs and I would not want to be involved." *(suspicious)*
Segment 3		
Aaron	"I would respect what he says and not share anything to anyone." *(go ahead/accept offer/comply with request)*	"I have a similar job right now that I can't tell people what projects I specifically work on." *(no reference to suspicion)*
Amy	"I would try to get the manager to help me sort this situation out." *(extricate)*	"Because the man could be lying but he also could be telling the truth." *(suspicious – indecisive)*
Jia	"I would not pay." *(decide against/decline offer/deny request)*	"I do not trust this course as it seems fishy." *(suspicious)*
Segment 4		
Derek	"I would sit around." *(continue as were)*	"If she wanted her stuff back she would message me" *(no reference to suspicion)*
Aaron	"I would be enraged and feel like an idiot. I would call the police immediately." *(involve authorities)*	"This man is a monster and he is now trying to make me an accomplice." *(suspicious)*
Amy	"Make the gentleman aware that I'm ready to serve him if he wants." *(take action)*	"I think it would be polite to make the customers aware that I am ready to serve" *(no reference to suspicion)*
Segment 5		
Jia	"seek help from authorities." *(involve authorities)*	"I would realize I have been scammed and would need help." *(suspicious)*
Andy	"I'd ask about what the job I am doing and who the people are I'm picking up." *(seek info/clarification)*	"I'm starting to believe that something is dodgy and that I don't really want to be a part of anything illegal." *(suspicious)*
Aaron	"I guess he could be telling the truth" *(go ahead/accept offer/comply with request)*	"I don't want any confrontation." *(no reference to suspicion)*
Segment 6		
Andy	"Say sure to the boy and wait around." *(go ahead/accept offer/comply with request)*	"The boy is vulnerable and I want to make sure that he is safe." *(no reference to suspicion)*

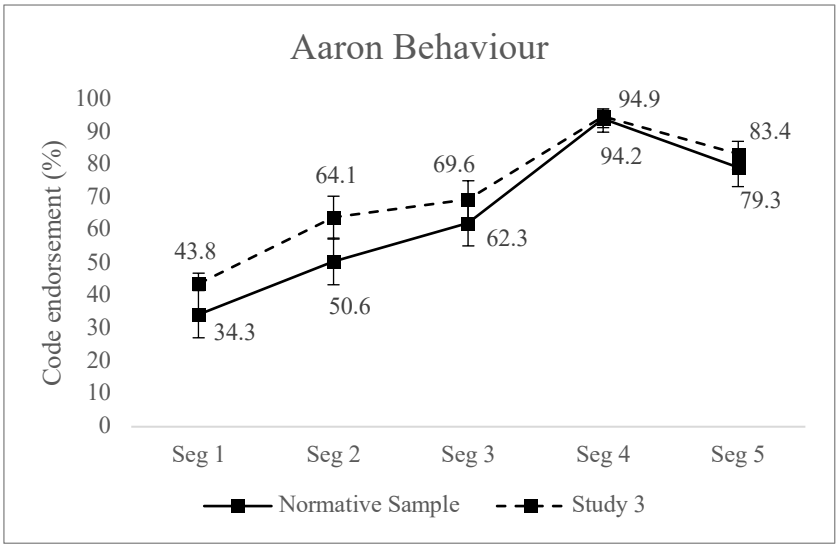
Significant chi-square outcomes are displayed in Table 19. The remaining non-significant chi-squares appear in Appendix Q (pp. 210-211). Although, as expected, the response patterns differed across scenarios, they generally indicated that adaptive behaviours and suspicion were uncommon at the first segment but highly likely by the final segment, with varying trends across the central segments. The Andy crime scenario continued to display mixed responding, being the only scenario to receive less than 50% adaptive behaviours and 65% suspicion by the final segment across both studies. In the non-crime scenario, although the descriptive statistics indicate a slight increase in suspicion, the percentage of overall suspicious responses remained below 40% at all segments, indicating that most participants were not reporting suspicion across all scenarios. Considering the 'adaptive' category contained responses that included seeking further information, it is unsurprising that the proportion of adaptive behaviours was high at some points. Figures 25a to 25e display the response patterns from the normative sample and Study 3. The correspondence between both the adaptive and suspicion data patterns for the two samples for each of the five scenarios is striking, with the 95% confidence interval error bars indicating substantial overlap between the patterns for the two samples.

Table 19
Significant chi-square outcomes for comparisons of SAP responses from the 'pseudo-normative' data and Study 3

	N	Chi Square	φ
Aaron			
Behaviour Segment 1	428	χ^2 (1) = 3.97, p =.046	.10
Segment 2	429	χ^2 (1) = 7.79, p=.005	.14
Suspicion Segment 2	442	χ^2 (1) = 11.58, p <.001	.16
Segment 3	445	χ^2 (1) = 4.59, p= .032	.10
Andy			
Behaviour Segment 2	442	χ^2 (1) = 5.14, p=.023	.11
Suspicion Segment 2	448	χ^2 (1) = 4.10, p =.043	.10
Segment 3	447	χ^2 (1) = 5.77, p =.016	.11
Amy			
Behaviour Segment 5	443	χ^2 (1) = 4.98, p =.026	-.11
Suspicion Segment 1	451	χ^2 (1) = 6.99, p =.008	.12

Table 19 (Continued)

Jia			
Behaviour Segment 1	431	$\chi^2(1) = 14.28, p < .001$	.18
Segment 3	439	$\chi^2(1) = 22.70, p < .001$	.23
Suspicion Segment 1	440	$\chi^2(1) = 8.93, p = .003$	.14
Segment 2	443	$\chi^2(1) = 4.50, p = .034$	.10
Segment 3	452	$\chi^2(1) = 11.45, p < .001$	.16
Derek			
Behaviour Segment 1	449	$\chi^2(1) = 4.90, p = .027$	.10
Segment 3	447	$\chi^2(1) = 3.93, p = .047$	.09
Suspicion Segment 2	452	$\chi^2(1) = 9.46, p = .002$	.15



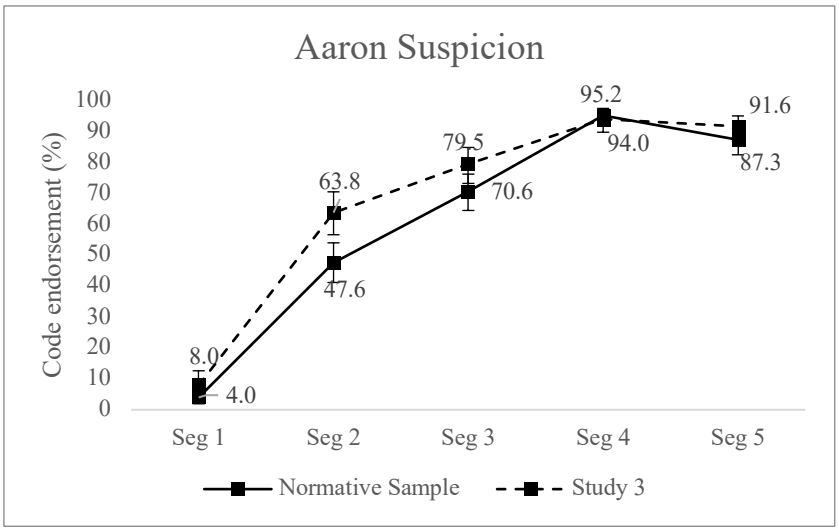


Figure 25a. *Percentage of adaptive behaviours and suspicion responses in the Aaron scenario for the 'pseudo-normative' data (Study 1 and 2) and Study 3, with 95% CI error bars*.

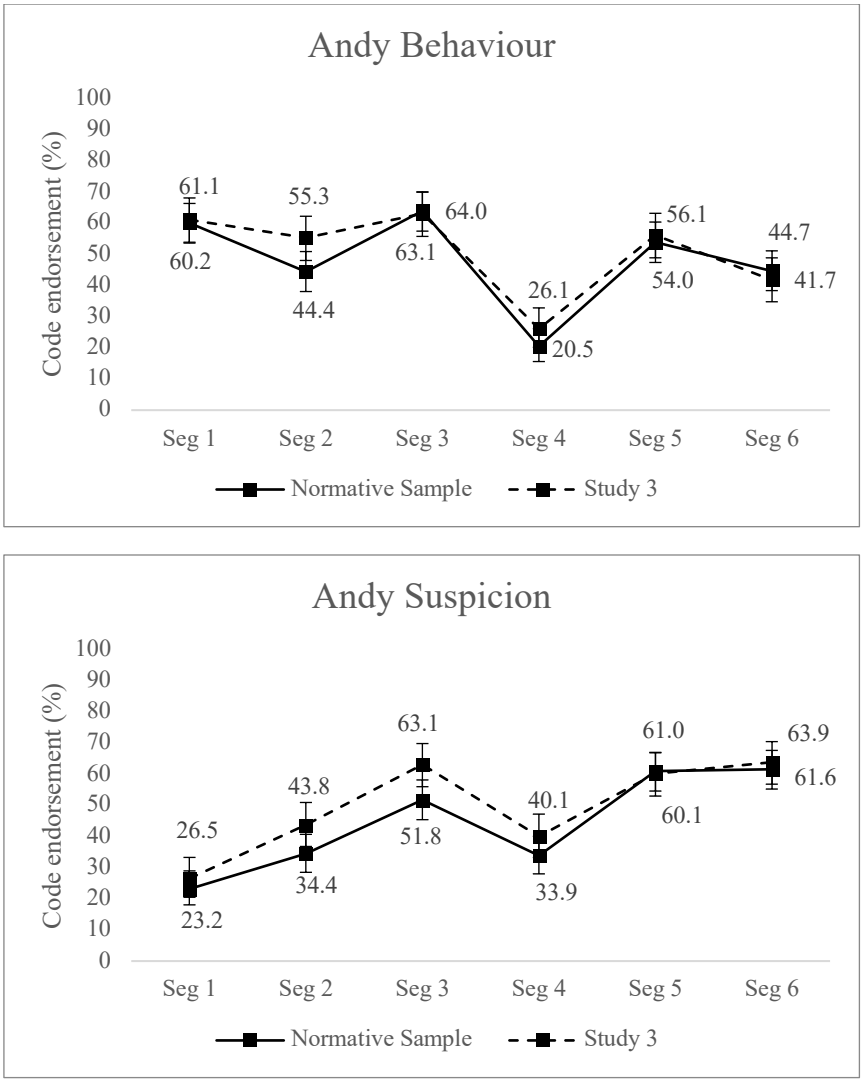
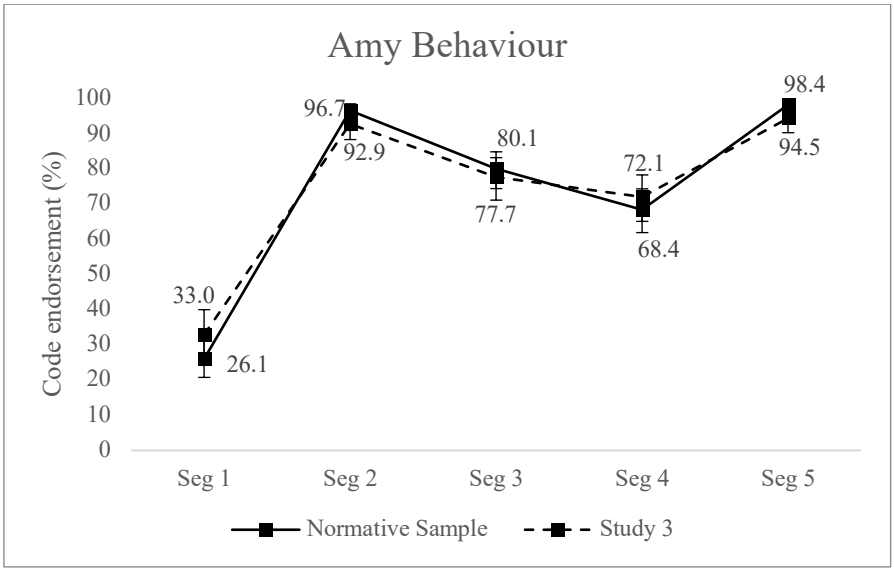


Figure 25b. *Percentage of adaptive behaviours and suspicion responses in the Andy scenario for the normative data (Study 1 and 2) and Study 3, with 95% CI error bars*.



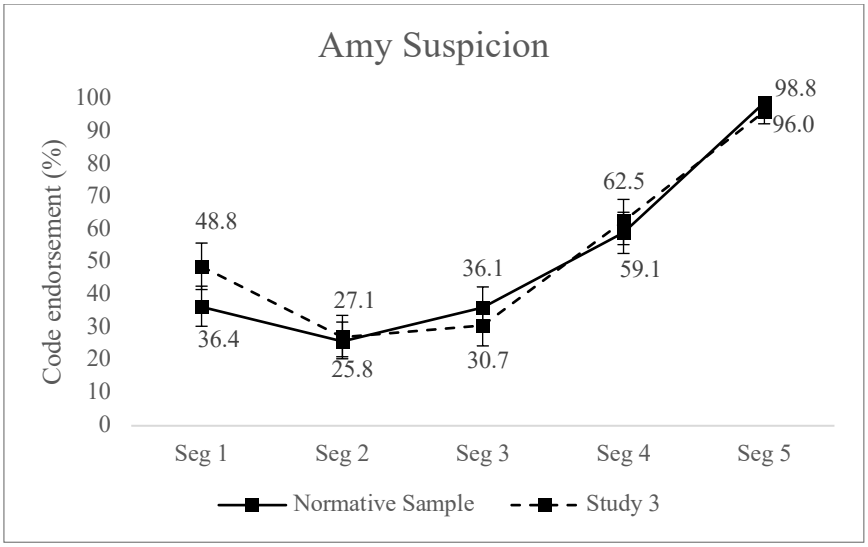
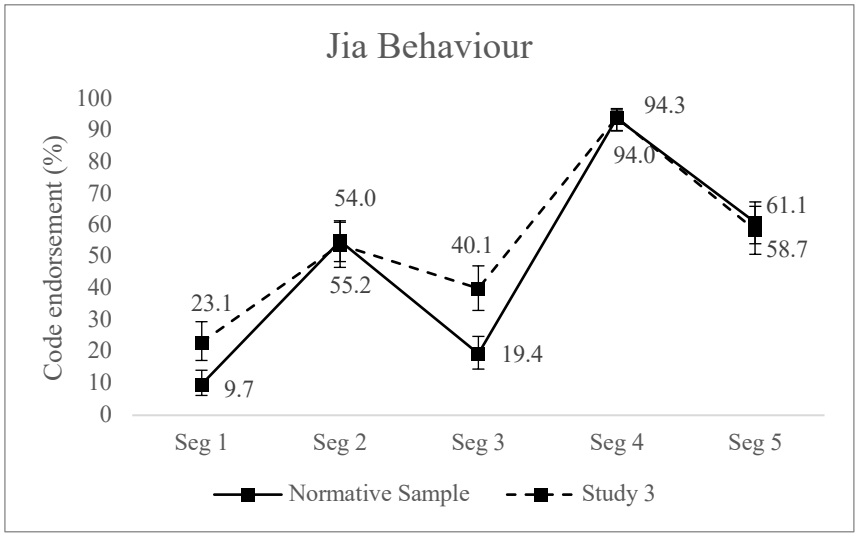


Figure 25c. *Percentage of adaptive behaviours and suspicion responses in the Amy scenario for the normative data (Study 1 and 2) and Study 3, with 95% CI error bars*.



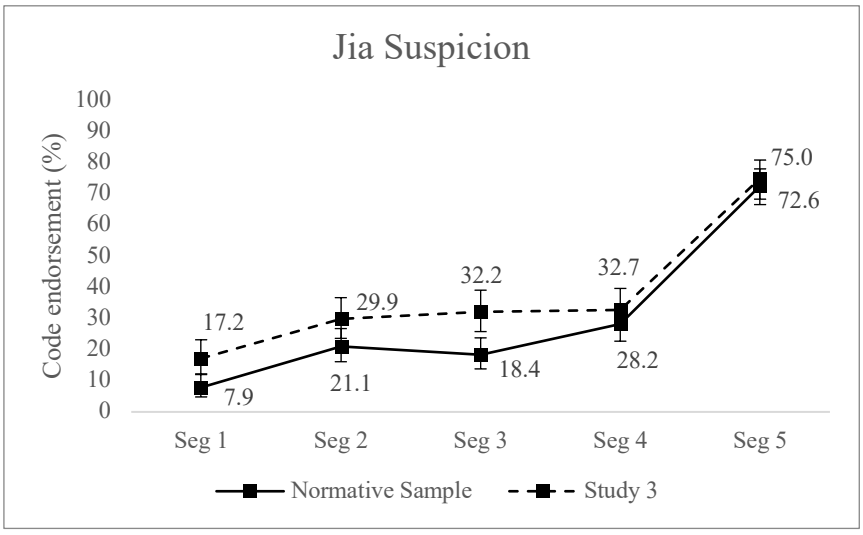
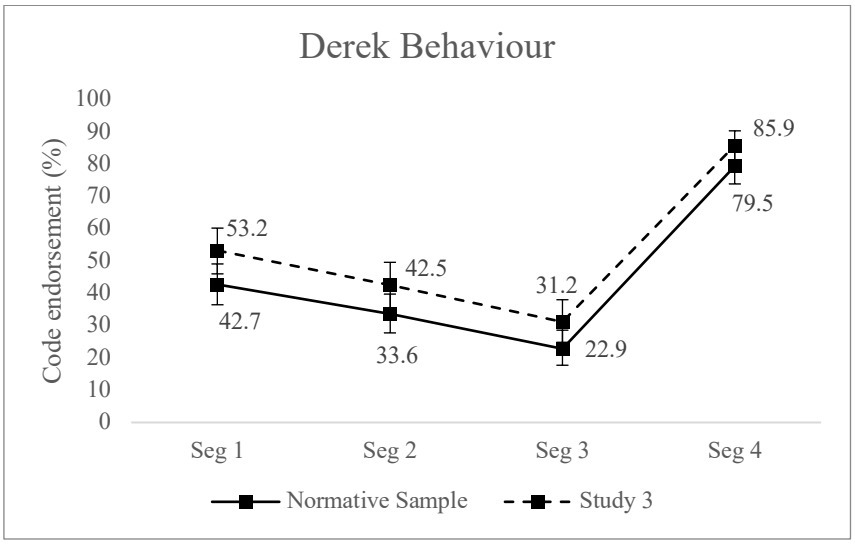


Figure 25d. *Percentage of adaptive behaviours and suspicion responses in the Jia scenario for the normative data (Study 1 and 2) and Study 3, with 95% CI error bars*.



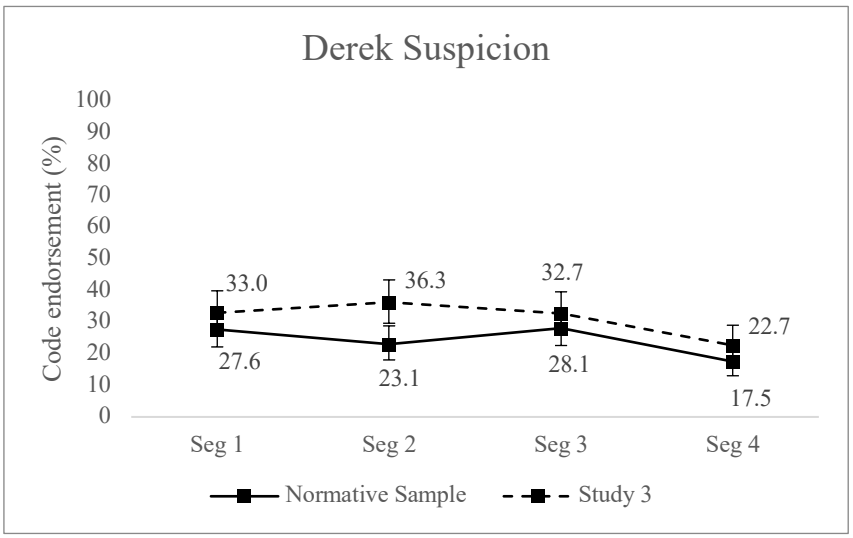


Figure 25e. *Percentage of adaptive behaviours and suspicion responses in the Derek scenario for the normative data (Study 1 and 2) and Study 3, with 95% CI error bars*.

Comparison of Autistic and Non-autistic Adults' SAP Responses

To examine whether autistic individuals were less likely to report adaptive behaviours and suspicion than non-autistic individuals, responses from autistic and non-autistic participants were compared. The response patterns revealed minimal variations between the two groups (see Figures 26a to 26e). In fact, at no point did the autistic participants demonstrate less suspicion or fewer adaptive behaviours than non-autistic participants. Chi-squares were run to identify any significant differences. They revealed only four (from a total

of 25) segments at which the groups differed; in each case autistic participants responded with more adaptive behaviours or suspicion than non-autistic participants: Andy segment 2 suspicion: $\chi^2(1) = 4.37$, p=.037, φ = .15, Amy segment 3 suspicion: $\chi^2(1) = 6.05$, p=.014, φ = .18, Jia segment 2 behaviour: $\chi^2(1) = 4.15$, p=.042, φ = .15, and Derek segment 4 suspicion: $\chi^2(1) = 4.37$, p=.037, φ = .15. Notably, the effect sizes are small, and the 95% CI error bars reinforce the observation that patterns for the two groups are similar[5].

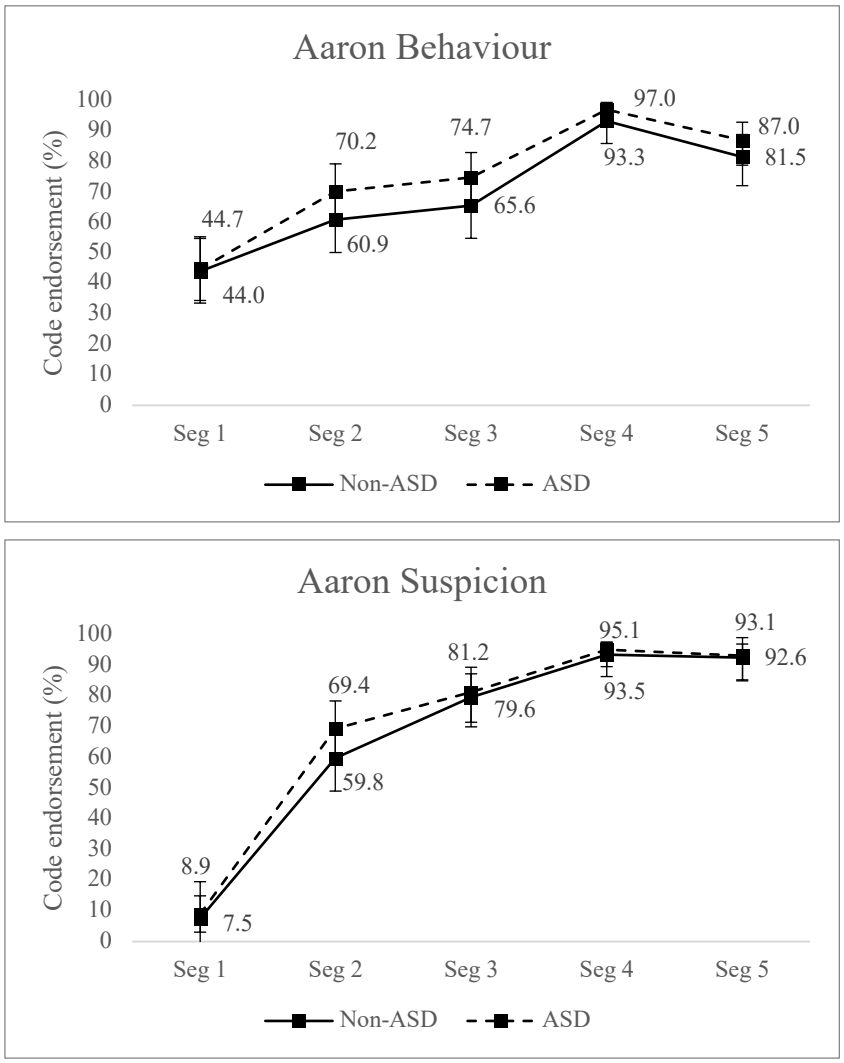
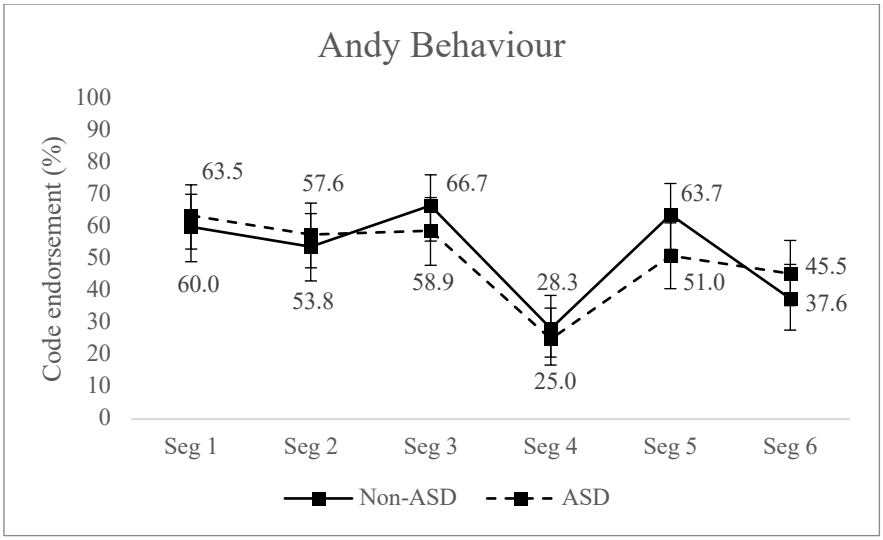


Figure 26a. *Percentage of adaptive behaviour and suspicion responses in the Aaron scenario for the autistic and non-autistic groups, with 95% CI error bars*.

[5] Given the marginally higher verbal ability of the autistic sample and the previously reported association between VCI and SAP performance, analyses were re-run after trimming the samples to exclude those autistic and non-autistic participants whose VCIs were higher and lower, respectively, than the upper and lower values of the other group.. The resulting patterns matched those displayed here.



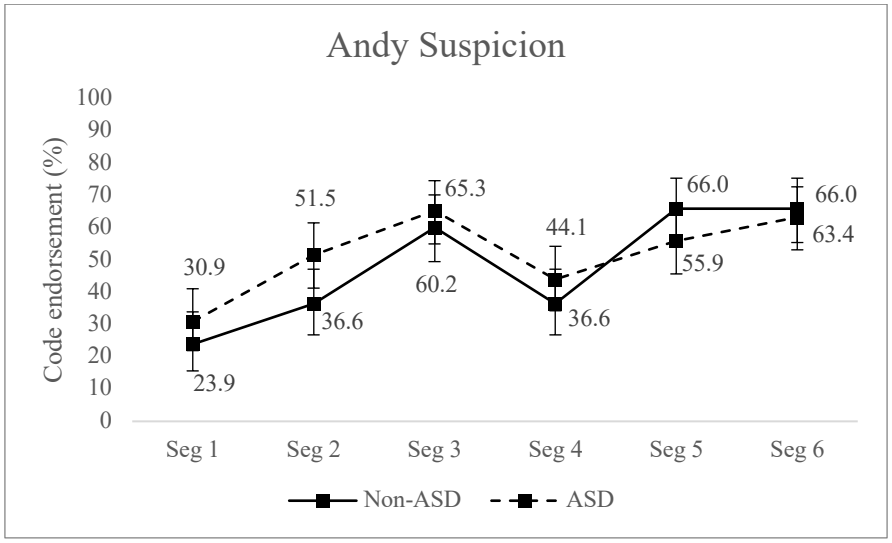
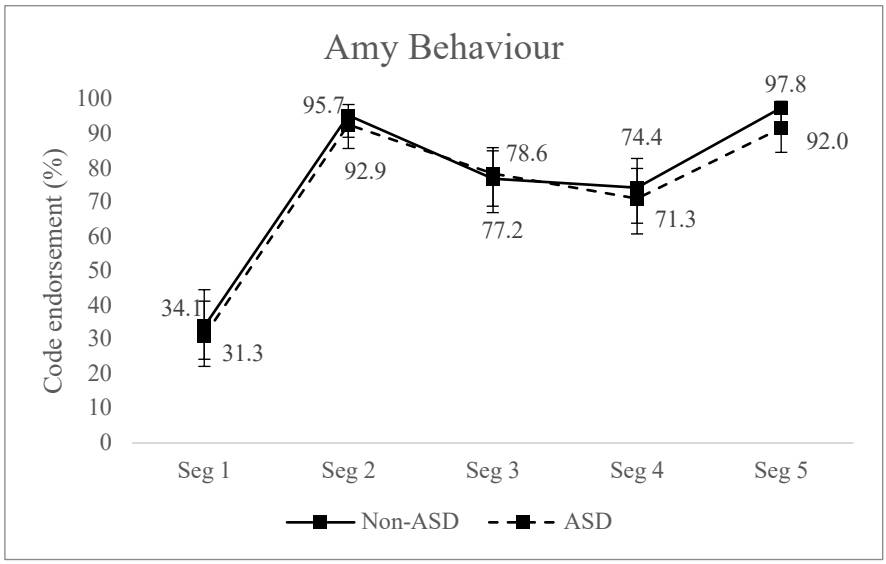


Figure 26b. *Percentage of adaptive behaviour and suspicion responses in the Andy scenario for the autistic and non-autistic groups, with 95% CI error bars*.



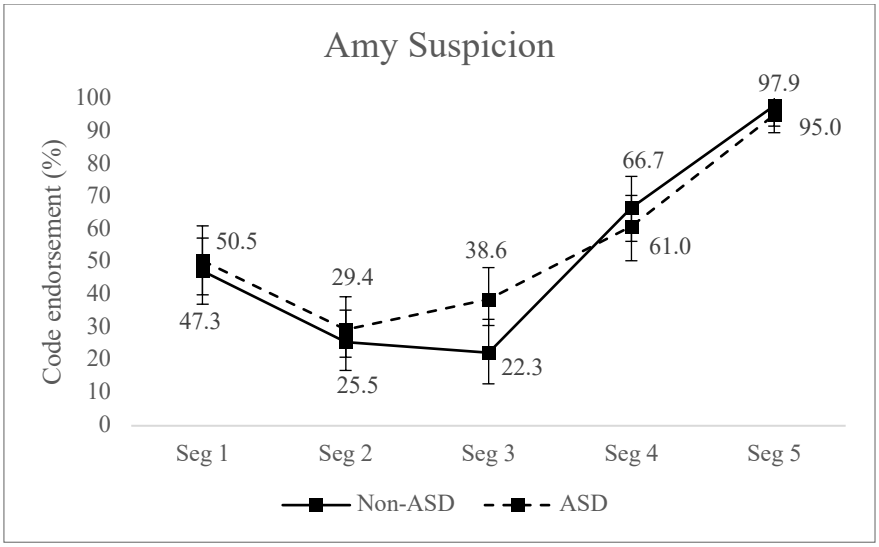
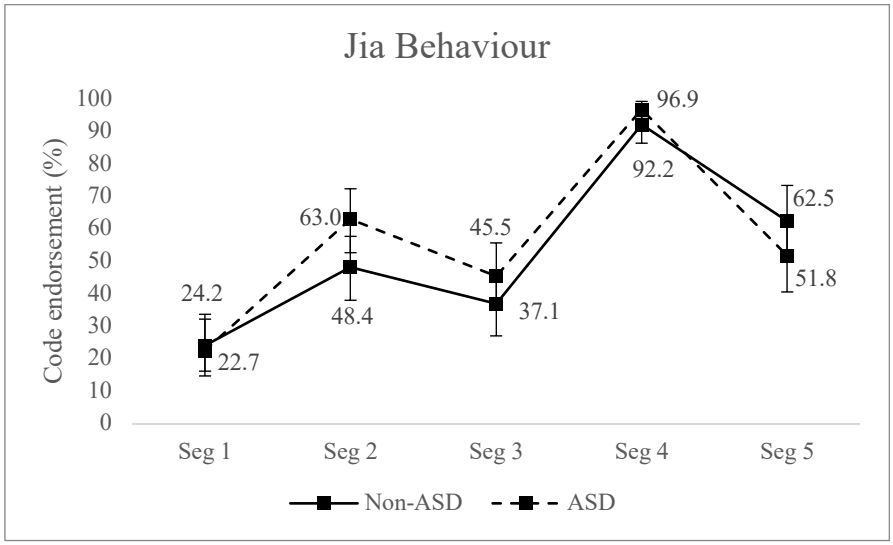


Figure 26c. *Percentage of extricating behaviour and suspicion responses in the Amy scenario for the autistic and non-autistic groups, with 95% CI error bars*.



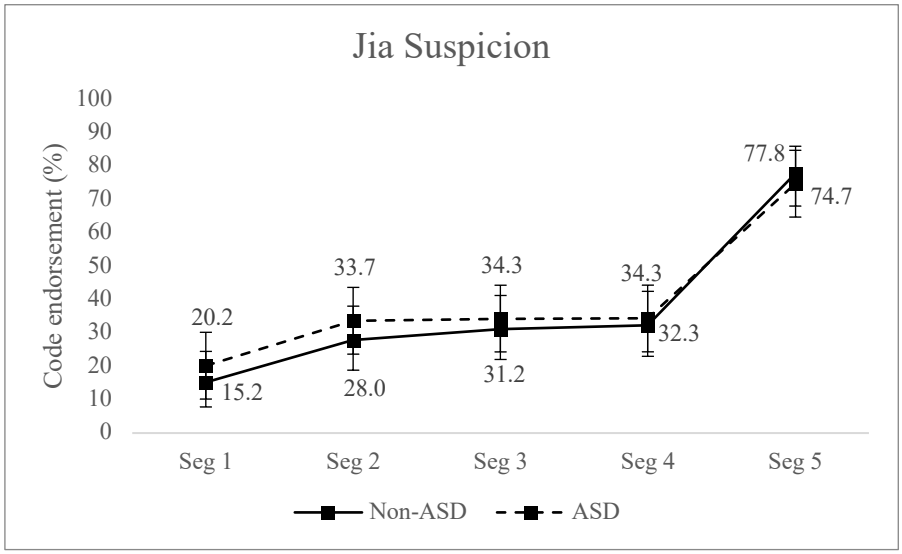
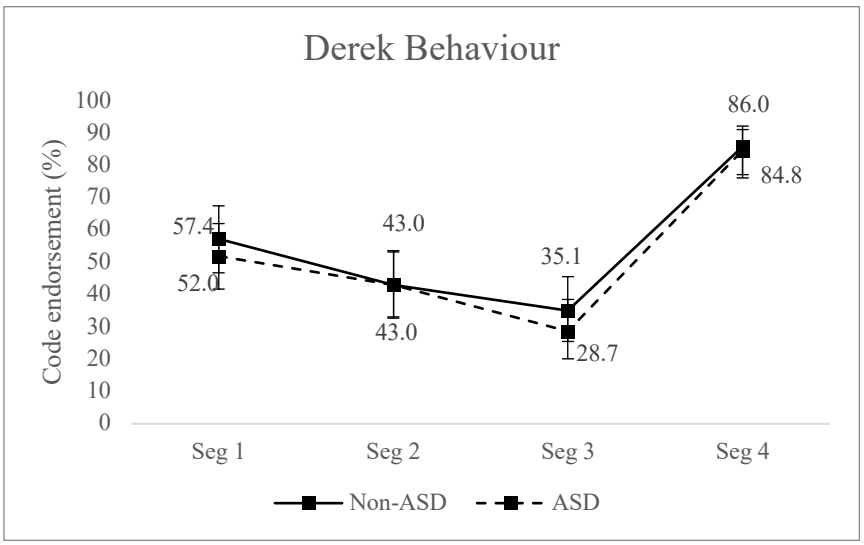


Figure 26d. *Percentage of adaptive behaviour and suspicion responses in the Jia scenario for the autistic and non-autistic groups, with 95% CI error bars*.



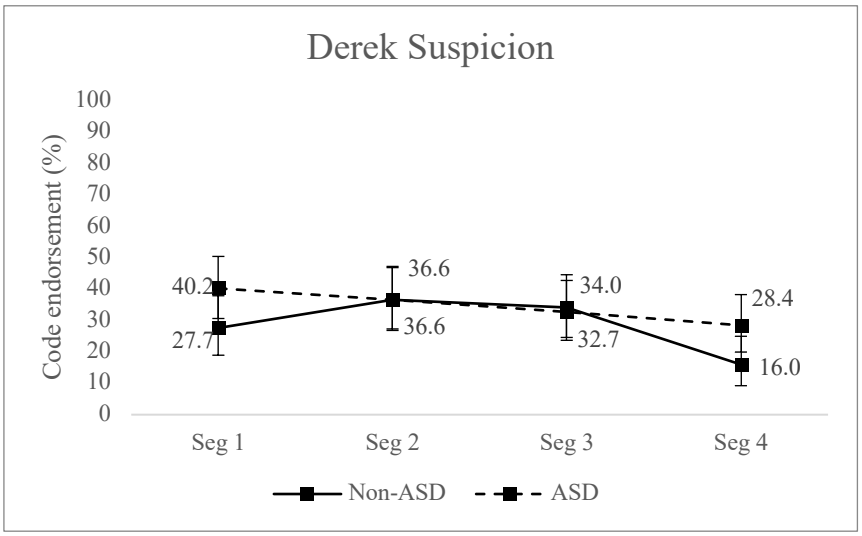


Figure 26e. *Percentage of adaptive behaviour and suspicion responses in the Derek scenario for the autistic and non-autistic groups, with 95% CI error bars*.

Comparison of Autistic and Non-autistic Adults' SAP Confidence Ratings

To investigate the possibility that SAP responses might have been related to participants' confidence, and whether this differed between the two groups, the 'sureness' ratings associated with each segment were examined. Frequency histograms revealed that confidence scores were highly negatively skewed and clustered near ceiling across the segments, indicating that participants were mostly confident in their responses. Group

comparisons indicated there were generally no significant differences between autistic and non-autistic participants' confidence (see Appendix R [p. 212-215] for t tests and Mann Whitney U tests). Furthermore, correlations between confidence and behaviour/suspicion were generally negligible to very weak, with 22/25 behaviour and 23/25 suspicion coefficients non-significant for the non-autistic group and 23/25 for both behaviour and suspicion in the autistic group (see Appendix S, pp. 216-217).

Relationship Between Theory of Mind and SAP Performance

Although a significant difference in ToM was revealed between the autistic and non-autistic participants, the autistic sample scored higher than reported in previous samples using the A-ToM-Q (Brewer et al., 2022; Brewer et al., 2023) and—as highlighted by the median scores reported in the Participants section—a substantial proportion of individuals in both groups scored at or near ceiling. Further, only 7.8% of all participants scored below 3 on the 6-point scale. Therefore, irrespective of diagnosis, it became apparent that the likelihood of detecting meaningful relationships between ToM and SAP performance would be constrained by (a) the range restriction on the A-ToM-Q and, relatedly (b) the relatively few individuals with low ToM scores who would be the ones most likely to perform poorly on the SAP. Nevertheless, to investigate whether ToM was associated with SAP performance, in the first instance I conducted Spearman's Rho point-biserial correlations due to the significant negative skew (-1.15) of the ToM data (see Table 20).

Compared with the previous study in which only two significant relationships emerged in the Andy scenario, several significant, though weak, relationships between ToM and SAP performance were revealed in the Aaron and Andy scenarios. In all but one case (Aaron segment 1 behaviour, ρ = -0.17, p =.022), higher ToM was associated with more adaptive behaviours and suspicion. An additional relationship was revealed in the suspicion reported at segment 1 in the Jia scenario, with higher ToM associated with more suspicion (ρ = 0.17, p

=.017). In the Amy crime scenario, neither behaviour nor suspicion were associated with ToM at any point. And, with the exception of segment 1 suspicion, neither behaviour nor suspicion were associated with ToM in the Jia scenario. No significant relationships were detected in the Derek non-crime scenario.

Table 20

Point biserial correlation coefficients (Spearman's Rho) for the SAP and A-ToM-Q for the four crime scenarios and the non-crime scenario

	Behaviour			Suspicion		
Segment	Rho	*p*	95% CI	Rho	*p*	95% CI
Aaron (crime)						
Segment 1	-.17	.022	-.31, -.02	.06	.392	-.08, .21
Segment 2	.09	.243	-.06, .23	.18	.013	.04, .32
Segment 3	.20	.006	.05, .34	.15	.042	.00, .29
Segment 4	-.13	.072	-.28, .02	-.07	.308	-.22, .07
Segment 5	.05	.461	-.09, .20	.04	.611	-.11, .18
Andy (crime)						
Segment 1	.12	.112	-.03, .26	.07	.368	-.08, .21
Segment 2	.15	.032	.01, .29	.14	.044	.00, .28
Segment 3	.17	.021	.02, .31	.14	.051	.00, .28
Segment 4	-.01	.885	-.16, .13	.10	.175	-.05, .24
Segment 5	.30	.000	.16, .43	.26	.000	.13, .39
Segment 6	.17	.020	.02, .30	.05	.449	-.09, .20
Amy (crime)						
Segment 1	.03	.725	-.12, .17	.06	.411	-.09, .20
Segment 2	-.05	.486	-.20, .10	-.08	.279	-.22, .07
Segment 3	-.04	.570	-.19, .11	.00	.994	-.15, .15
Segment 4	.00	.979	-.15, .15	.02	.803	-.13, .16
Segment 5	.05	.472	-.09, .20	-.01	.904	-.15, .14
Jia (crime)						
Segment 1	.04	.553	-.11, .19	.17	.017	.03, .31
Segment 2	.07	.311	-.07, .22	.12	.112	-.03, .26
Segment 3	.09	.232	-.06, .23	.09	.219	-.06, .23
Segment 4	-.01	.874	-.16, .14	.07	.311	-.07, .22
Segment 5	-.02	.764	-.18, .14	.09	.225	-.06, .23
Derek (non-crime)						
Segment 1	.10	.159	-.04, .25	-.02	.753	-.17, .12
Segment 2	-.01	.941	-.15, .14	.02	.832	-.13, .16
Segment 3	-.08	.262	-.23, .07	.01	.914	-.14, .15
Segment 4	.05	.519	-.10, .19	.07	.305	-.07, .22

SAP performance of high and low ToM scorers. Although the distributional characteristics of the ToM data for the entire sample mitigated against detecting meaningful ToM–SAP relationships, an alternative—albeit post hoc—way of examining the data provided a promising indication of a relationship, highlighting that individuals who scored very poorly on the A-ToM-Q were both less likely to respond adaptively, and to do so at an early rather than a later segment, than those who scored highly.

For this analysis, each participant was first assigned a score ranging from 0-4 for each of the segments 1 to 5; this score reflected the number of times they responded adaptively (or reported suspicion) at that segment across all of the crime scenarios. For example, a person who responded adaptively three out of four times at segment 1 across the four crime scenarios was assigned a score of three for that variable. Although the Andy crime scenario contained six segments, only the first five were included. Participants who were previously excluded from analyses at each segment across any of the crime scenarios (due to audio or technical failure, or a response that fell into the 'non-behavioural' or 'other' category), were excluded from the analyses. That is, if a person had three valid scores and one missing data value at segment 2, they were not included in the analyses for that segment.

It is obviously somewhat arbitrary as to what constitutes high and low A-ToM-Q scorers. To highlight the relationship between extreme ToM scores and adaptive responding, I present data that illustrate the difference in responding between high and low ToM scorers when the 'high ToM' group included participants with an A-ToM-Q score ≥ 5 ($n = 128$), and the 'low ToM' group included participants with an A-ToM-Q score ≤ 3 ($n = 30$). The lower cut-off was chosen because too few participants (only 7.8% of the total sample) scored ≤ 2, the next possible cutoff. Participants with an A-ToM-Q score of four ($n = 35$) were excluded from the analyses to provide a clearer delineation between sub-groups. The two sub-groups did not differ significantly in verbal ability (low ToM: $M = 102.33$, $SD = 7.27$, $Mdn = 101.0$,

high ToM: M = 104.66, SD = 8.37, Mdn = 104.0), t (156) = -1.40, p =.163, d = -.28, 95% CI [-.68, .12], U = 2262.00, p =.129.

Examination of the frequency data revealed (see Table 21) that, at the second segment, only 38.5% of low ToM scorers reported adaptive behaviours more than twice across the four crime scenarios, compared with 59.8% of high ToM scorers. At segment 3, 61.8% of the high ToM group reported more than two adaptive behaviours compared with 44% of the low ToM group, and by segment 5 the difference had increased to 78.4% for the high ToM group compared with 45.5% for the low ToM group. Similar trends emerged for the suspicion data. At segment 3, 44.3% of the high ToM group had reported suspicion more than twice compared with only 17.8% of the low ToM group, and by the fifth segment 90.3% of the high ToM group reported suspicion more than twice, compared with 67.9% of the low ToM group. This trend is highlighted by plots of the cumulative percentage of adaptive behaviours and suspicion at each segment (see Figure 27a and 27b and Appendix T, pp. 218-220) which emphasise that the low ToM group were less likely to offer adaptive responses and indicate suspicion across the various crime scenarios.

Table 21
Observed frequencies (and percentages) for the number of adaptive behaviours by high (≥ 5) versus low (≤ 3) ToM scorers

No. of responses	Behaviour		Suspicion	
	Low ToM	High ToM	Low ToM	High ToM
	n (%)	*n (%)*	*n (%)*	*n (%)*
Segment 1				
0	4 (18.2)	15 (13.4)	11 (42.3)	40 (33.3)
1	8 (36.4)	37 (33.0)	12 (46.2)	37 (30.8)
2	7 (31.8)	34 (30.4)	1 (3.8)	27 (22.5)
3	2 (9.1)	20 (17.9)	2 (7.7)	16 (13.3)
4	1 (4.5)	6 (5.4)	0 (0)	0 (0)
Segment 2				
0	0 (0)	1 (0.9)	6 (22.2)	16 (13.1)
1	5 (19.2)	9 (8.0)	10 (37.0)	36 (29.5)
2	11 (42.3)	35 (31.3)	8 (29.6)	37 (30.3)

Table 21 (Continued)

3	6 (23.1)	36 (32.1)	2 (7.4)	23 (18.9)
4	4 (15.4)	31 (27.7)	1 (3.7)	10 (8.2)
Segment 3				
0	1 (4.0)	2 (1.8)	3 (10.7)	6 (4.9)
1	7 (28.0)	15 (13.6)	6 (21.4)	29 (23.8)
2	6 (24.0)	25 (22.7)	14 (50.0)	33 (27.0)
3	9 (36.0)	44 (40.0)	3 (10.7)	39 (32.0)
4	2 (8.0)	24 (21.8)	2 (7.1)	15 (12.3)
Segment 4				
0	0 (0)	0 (0)	2 (6.9)	4 (3.3)
1	0 (0)	7 (6.4)	7 (24.1)	19 (15.4)
2	8 (32.0)	17 (15.6)	11 (37.9)	46 (37.4)
3	15 (60.0)	61 (56.0)	7 (24.1)	31 (25.2)
4	2 (8.0)	24 (22.0)	2 (6.9)	23 (18.7)
Segment 5				
0	1 (4.5)	1 (1.0)	0 (0)	3 (2.4)
1	4 (18.2)	2 (2.0)	2 (7.1)	0 (0)
2	7 (31.8)	19 (18.6)	7 (25.0)	9 (7.3)
3	6 (27.3)	45 (44.1)	12 (42.9)	43 (35.0)
4	4 (18.2)	35 (34.3)	7 (25.0)	68 (55.3)

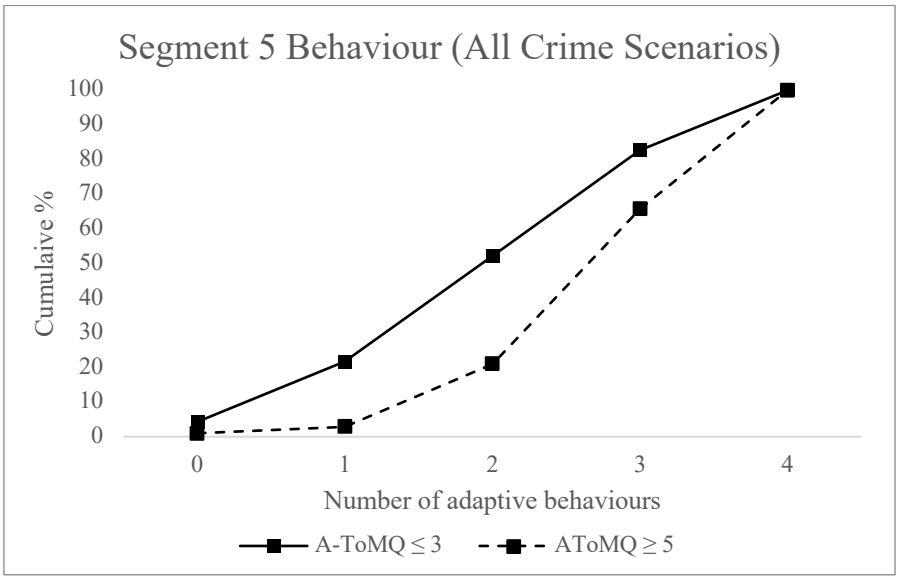


Figure 27a. *Cumulative percentages for the number of adaptive behaviours reported in Segment 5 across the four crime scenarios*.

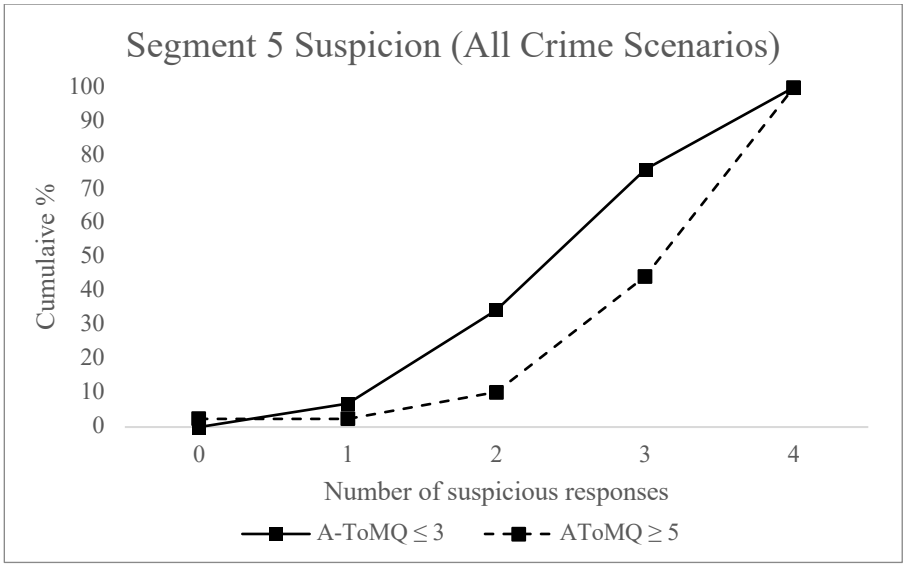


Figure 27b. *Cumulative percentages for the number of suspicion responses reported in Segment 5 across the four crime scenarios*.

To assess whether the high and low ToM sub-groups differed in the number of adaptive responses and suspicion, chi-square analyses were run for behaviour and suspicion at each segment. Due to an inadequate sample size for the chi-square test of homogeneity (Cochran, 1954), the Fisher-Freeman-Halton exact test (Freeman & Halton, 1951) was conducted. Analyses revealed that the response distributions differed significantly in both behaviour (p =.005) and suspicion (p <.001) at segment 5, and in suspicion at segment 3 (p =.042), indicating that the high ToM group were more likely to respond adaptively by the end of the crime scenario, and more likely to report suspicion earlier. See Table 22 for all Fisher-Freeman-Halton tests. An alternative way to analyse these data is to conduct a Mann-Whitney-U test which compares the ranks of the dependent variable (number of adaptive /suspicious responses) for the high and low ToM groups. The results of this analysis reveal similar patterns (see Table 23).

For further post-hoc comparisons, the number of adaptive or suspicious responses were split into two categories, with each participant categorised as having responded adaptively either twice or less, or more than twice (i.e., 3 or 4 times), and high versus low ToM scorers

Table 22

Fisher-Freeman-Halton tests comparing the high and low ToM groups on the number of adaptive and suspicious responses

Combined Crime	N	Fisher's Exact	φ
Behaviour			
Segment 1	134	χ^2 (4) = 1.37, p =.884	.10
Segment 2	138	χ^2 (4) = 5.48, p =.237	.20
Segment 3	135	χ^2 (4) = 5.48, p =.216	.20
Segment 4	134	χ^2 (4) = 5.87, p =.102	.22
Segment 5	124	χ^2 (4) = 13.37, p =.005	.35
Suspicion			
Segment 1	146	χ^2 (4) = 6.68, p =.077	.21
Segment 2	149	χ^2 (4) = 3.74, p =.436	.16
Segment 3	150	χ^2 (4) = 9.37, p =.042	.25
Segment 4	152	χ^2 (4) = 4.16, p =.369	.16
Segment 5	151	χ^2 (4) = 16.77, p <.001	.37

Table 23

Descriptive statistics and Mann Whitney U tests for high and low ToM group comparison on the number of times adaptive behaviours or suspicion were reported in the crime scenarios combined.

	N*	Mean Rank (Median)		*U*	*p*
		A-ToM ≤ 3	A-ToM ≥ 5		
Behaviour					
Segment 1	22 (112)	60.68 (1)	68.84 (2)	1382.00	.349
Segment 2	26 (112)	55.58 (2)	72.73 (3)	1818.00	.040
Segment 3	25 (110)	53.66 (2)	71.26 (3)	1733.50	.034
Segment 4	25 (109)	58.42 (3)	69.58 (3)	1589.50	.148
Segment 5	22 (102)	43.36 (2)	66.63 (3)	1543.00	.004
Suspicion					
Segment 1	26 (120)	61.06 (1)	76.20 (1)	1883.50	.083
Segment 2	27 (122)	61.04 (1)	78.09 (2)	2024.00	.055
Segment 3	28 (122)	62.52 (2)	78.48 (2)	2071.50	.069
Segment 4	29 (123)	63.93 (2)	79.46 (2)	2148.00	.075
Segment 5	28 (123)	53.20 (3)	81.19 (4)	2360.50	<.001

* Number of participants displayed as N = A-ToM ≤ 3(A-ToM ≥ 5)

were compared with 2 × 2 chi-square tests of homogeneity and Fisher-Freeman-Halton exact tests. The results presented vary depending on whether the expected counts were sufficient to meet chi-square test assumptions; therefore, the Fisher exact test outcome is displayed

without a test statistic. Like the earlier group comparisons, at segments 2, $\chi 2$ (1) = 3.90, p=.048, and 5, $\chi 2$ (1) = 3.90, p=.048, the high ToM group were more likely than the low ToM group to respond adaptively more than twice. Similarly, the high ToM group were more likely than the low ToM group to report suspicion more than twice at segment 3, $\chi 2$ (1) = 6.65, p=.010, and segment 5, p =.005. See Table 24 for all descriptive statistics and Table 25 for remaining chi-square statistics.

Table 24
Observed frequencies (and percentages) of ToM group and number of adaptive responses (≤ 2 or > 2)

No. of responses	Behaviour		Suspicion	
	Low ToM	High ToM	Low ToM	High ToM
	n (%)	*n (%)*	*n (%)*	*n (%)*
Segment 1				
≤ 2	19 (86.4)	86 (76.8)	24 (92.3)	22 (86.7)
> 2	3 (13.6)	26 (23.2)	10 (7.7)	80 (13.3)
Segment 2				
≤ 2	16 (61.5)	45 (40.2)	24 (88.9)	89 (73.0)
> 2	10 (38.5)	67 (59.8)	3 (11.1)	33 (27.0)
Segment 3				
≤ 2	14 (56.0)	42 (38.2)	23 (82.1)	68 (55.7)
> 2	11 (44.0)	68 (61.8)	5 (17.9)	54 (44.3)
Segment 4				
≤ 2	8 (32.0)	24 (22.0)	20 (69.0)	69 (56.1)
> 2	17 (68.0)	85 (78.0)	9 (31.0)	54 (43.9)
Segment 5				
≤ 2	12 (54.5)	22 (21.6)	9 (32.1)	12 (9.8)
> 2	10 (45.5)	80 (78.4)	19 (67.9)	111 (90.2)

SAP Performance and Autistic Traits

As noted when describing the samples, the autistic group reported higher autistic traits than the non-autistic group. Autistic traits were also correlated with SAP performance at several points in the scenarios. Contrary to the perspective that autistic individuals might be less likely to detect dodgy or suspicious behaviour, higher levels of autistic traits were associated with greater reporting of adaptive behaviours and suspicion in all cases where an

Table 25
Fisher-Freeman-Halton exact tests comparing high and low ToM group by response number group (≤ 2 or > 2)

≤ 2 or > 2	N	Chi-square or Fisher's exact test	φ
Behaviour			
Segment 1	134	p =.406	
Segment 2	138	χ^2 (1) = 3.90, p=.048	.17
Segment 3	135	χ^2 (1) = 2.66, p=.103	.14
Segment 4	134	χ^2 (1) = 1.12, p =.291	.09
Segment 5	124	χ^2 (1) = 9.89, p =.002	.28
Suspicion			
Segment 1	146	p =.742	
Segment 2	149	χ^2 (1) = 3.07, p=.080	.14
Segment 3	150	χ^2 (1) = 6.65, p=.010	.21
Segment 4	152	χ^2 (1) = 1.60, p =.206	.10
Segment 5	151	p =.005	

association was found. For example, in the Aaron scenario, autistic traits were positively associated with adaptive behaviours at segments 2, ρ = 0.26, p <.001, and 5, ρ = 0.16, p =.024, and suspicion at segments 2, ρ = 0.20, p =.005, and 3, ρ = 0.14, p =.043. In the Andy scenario, autistic traits were positively associated with suspicion at segments 1, ρ = 0.15, p =.038, and 2, ρ = 0.15, p =.034, and in the Jia scenario, autistic traits were positively associated with adaptive behaviours at segment 2, ρ = 0.21, p =.003, and 3, ρ = 0.23, p =.001, and suspicion at segment 2, ρ = 0.15, p =.038, and 3, ρ = 0.15, p =.034. In the non-crime Derek scenario, autistic traits were positively associated with suspicion at segment 1, ρ = 0.15, p =.038. There was no evidence of an association between autistic traits and behaviour or suspicion in the Amy scenario. The point biserial correlation figures for each scenario are displayed in Table 26. Note that most of the above patterns are consistent with the slightly higher levels of reporting of adaptive behaviours and suspicion by the autistic group as indicated by the curves depicted in Figures 26a to 26e.

Table 26

Point biserial correlation coefficients (Spearman's Rho) for the SAP and AQ-12

	Behaviour			Suspicion		
Segment	Rho	*p*	95% CI	Rho	*p*	95% CI
Aaron						
Segment 1	-.06	.403	-.20, .09	.03	.669	-.11, 0.17
Segment 2	.26	<.001	.12, .39	.20	.005	.06, 0.34
Segment 3	.11	.145	-.04, .25	.14	.043	.00, 0.28
Segment 4	.12	.104	-.03, .26	.11	.125	-.03, 0.25
Segment 5	.16	.024	.02, .30	.12	.092	-.02, 0.26
Andy						
Segment 1	.10	.148	-.04, .25	.15	.038	.00, 0.29
Segment 2	.10	.153	-.04, .24	.15	.034	.01, 0.29
Segment 3	-.10	.175	-.24, .05	.06	.426	-.09, 0.20
Segment 4	.05	.497	-.10, .19	.05	.452	-.09, 0.19
Segment 5	.03	.629	-.11, .18	.09	.180	-.05, 0.23
Segment 6	.06	.417	-.09, .20	.05	.495	-.09, 0.19
Amy						
Segment 1	.00	.960	-.15, .14	-.07	.337	-.21, 0.08
Segment 2	.07	.307	-.07, .21	.04	.604	-.11, 0.18
Segment 3	-.06	.443	-.20, .09	.13	.072	-.02, 0.26
Segment 4	.10	.169	-.05, .24	.05	.490	-.09, 0.19
Segment 5	-.05	.450	-.20, .09	.10	.177	-.05, 0.23
Jia						
Segment 1	.08	.252	-.06, .22	.11	.118	-.03, 0.25
Segment 2	.21	.003	.07, .34	.18	.009	.04, 0.32
Segment 3	.23	.001	.08, .36	.14	.042	.00, 0.28
Segment 4	.03	.646	-.11, .18	.13	.060	-.01, 0.27
Segment 5	-.08	.308	-.23, .08	.01	.916	-.14, 0.15
Derek						
Segment 1	.03	.688	-.11, .17	.15	.038	.00, 0.28
Segment 2	.06	.403	-.08, .20	.06	.408	-.08, 0.20
Segment 3	.06	.370	-.08, .20	.08	.235	-.06, 0.22
Segment 4	-.04	.570	-.18, .10	-.01	.928	-.15, .14

Relationship Between Verbal Ability and SAP Performance

Given (a) Brewer et al. (2023) detected significant relationships (independent of ToM) between verbal ability and dodginess discrimination, and (b) several relationships between verbal ability and SAP performance emerged in Study 2, point-biserial correlations between verbal ability and SAP performance were again examined. The Pearson correlation is displayed due to relative normality of the Proxy VCI data (skewness = -.37, kurtosis = -.02).

Several significant relationships were detected across the Aaron, Andy, Jia, and Derek scenarios, with participants with higher verbal ability scores more likely to report adaptive behaviours and suspicion. One exception to this pattern occurred in the first segment of the Aaron scenario, where verbal ability was negatively correlated with adaptive behaviours, indicating that participants with higher verbal ability reported more maladaptive behaviours at that point (r_{pb} = -.16, p =.029). The point biserial correlations are displayed in Table 27.

Table 27

Point biserial correlation coefficients (Pearson coefficient) for the SAP and Proxy VCI

	Behaviour			Suspicion		
Segment	*r*	*p*	95% CI	*r*	*p*	95% CI
Aaron						
Segment 1	-.16	.029	-.30, -.02	.06	.424	-0.08, .20
Segment 2	.39	<.001	.26, .50	.36	<.001	0.23, .48
Segment 3	.25	.001	.11, .37	.28	<.001	0.14, .40
Segment 4	.12	.104	-.02, .26	.07	.317	-0.07, .21
Segment 5	.18	.014	.04, .31	.03	.697	-0.11, .17
Andy						
Segment 1	.24	.001	.09, .37	.18	.015	.03, .31
Segment 2	.25	<.001	.12, .38	.21	.003	.07, .34
Segment 3	-.07	.366	-.21, .08	.06	.400	-.08, .20
Segment 4	-.01	.893	-.15, .13	.05	.477	-.09, .19
Segment 5	.25	<.001	.11, .38	.26	<.001	.12, .38
Segment 6	.07	.315	-.07, .21	-.05	.512	-.19, .09
Amy						
Segment 1	-.02	.746	-.17, .12	-.04	.613	-.18, .10
Segment 2	.05	.465	-.09, .19	-.07	.362	-.20, .08
Segment 3	.05	.519	-.10, .19	.05	.508	-.09, .19
Segment 4	.01	.930	-.14, .15	.00	.978	-.14, .14
Segment 5	-.07	.326	-.21, .07	.03	.629	-.11, .17
Jia						
Segment 1	.14	.054	.00, .28	.24	.001	.10, .37
Segment 2	.31	<.001	.18, .43	.31	<.001	.17, .43
Segment 3	.30	<.001	.16, .42	.33	<.001	.19, .45
Segment 4	.11	.150	-.04, .25	.32	<.001	.19, .44
Segment 5	-.11	.170	-.26, .05	.10	.171	-.04, .24
Derek						
Segment 1	.28	<.001	.14, .40	.20	.005	.06, .33
Segment 2	.06	.433	-.09, .20	.17	.015	.03, .31
Segment 3	.09	.193	-.05, .23	.11	.122	-.03, .25
Segment 4	-.04	.565	-.18, .10	.18	.011	.04, .31

Relationship Between SAP Performance and Potential Concurrent and Divergent Validity Markers

First, I examined descriptive statistics for the autistic and non-autistic groups on the gullibility, social vulnerability and interpersonal trust measures. Group comparisons revealed that autistic participants (M = 38.27, SD = 16.62, Mdn = 34.00) reported higher levels of gullibility than non-autistic participants (M = 30.42, SD = 12.75, Mdn = 28.00), t (188.21) = -3.74, p<.001, d = -0.53, 95% CI [-.81, -.24], U = 6210.50, p <.001. There were no significant group differences in social vulnerability, which deviated from normality (autistic: M = 12.39, SD = 7.31, Mdn = 10.00; non-autistic: M = 10.25, SD = 5.73, Mdn = 10.00), U = 5590.50, p = .062), or interpersonal trust (autistic: M = 66.25, SD = 9.55, Mdn = 68.00; non-autistic: M = 67.71, SD = 10.21, Mdn = 67), t (195) = 1.03, p=.152, d = 0.15, 95% CI [-0.13, 0.43], U = 4729.00, p = .772. See Appendix U (p. 221) for these statistics in tabular form.

Evidence of concurrent validity for the SAP was sought by analysing the relationship between SAP performance and the potential concurrent markers of gullibility and social vulnerability, while evidence of divergent validity was expected to be revealed via null relationships between interpersonal trust and SAP responding. For gullibility, all correlations with SAP performance were negligible or very weak (see Table 28). Of the 50 correlation coefficients, only three were significant—an unsurprising outcome given the number of coefficients even if there is no meaningful relationship between the two variables—but higher self-reported gullibility was associated with more suspicion in Andy segments 1, ρ = .17, p =.032, and 2, ρ = .17, p =.029, and Jia segment 1, ρ = .17, p =.031.

These analyses were then repeated for SAP performance and social vulnerability (see Table 29). Again, all coefficients were negligible or very weak, with only 5 of the 50 coefficients significant. The significant coefficients indicated negative associations between social vulnerability and adaptive behaviours in Andy segment 2, ρ = -.15, p =.035, and 4, ρ =

-.16, p =.023, and Derek segment 1, ρ = -.15, p =.032. That is, as self-reported social vulnerability increased, the likelihood of somebody responding adaptively decreased. Significant negative associations were also found in Amy segment 2 suspicion, ρ = -.16, p =.021, and Jia segment 4 suspicion, ρ = -.15, p =.040.

Table 28

Point biserial correlation coefficients (Spearman's Rho) for the SAP and gullibility

	Behaviour			Suspicion		
Segment	*Rho*	*p*	95% CI	*Rho*	*p*	95% CI
Aaron						
Segment 1	.10	.286	-.09, .29	.02	.786	-.14, .18
Segment 2	.12	.207	-.07, .31	.07	.364	-.09, .23
Segment 3	.05	.637	-.15, .24	.06	.458	-.10, .22
Segment 4	.03	.727	-.16, .23	.13	.094	-.03, .29
Segment 5	.09	.374	-.11, .27	-.01	.895	-.17, .15
Andy						
Segment 1	.12	.230	-.08, .30	.17	.032	.01, .32
Segment 2	-.01	.886	-.21, .18	.17	.029	.01, .32
Segment 3	-.12	.203	-.31, .07	.06	.447	-.10, .22
Segment 4	-.01	.956	-.20, .19	.14	.073	-.02, .29
Segment 5	.08	.422	-.12, .27	.03	.704	-.13, .19
Segment 6	.08	.412	-.12, .27	.11	.157	-.05, .27
Amy						
Segment 1	-.05	.622	-.24, .15	.08	.295	-.08, .24
Segment 2	.10	.282	-.09, .29	.03	.673	-.13, .19
Segment 3	-.09	.341	-.28, .10	.01	.947	-.15, .16
Segment 4	.01	.894	-.18, .21	.07	.369	-.09, .23
Segment 5	.03	.719	-.16, .23	-.08	.338	-.23, .08
Jia						
Segment 1	.13	.178	-.07, .32	.17	.031	.01, .32
Segment 2	.02	.830	-.17, .21	.02	.847	-.14, .17
Segment 3	.12	.209	-.07, .31	.03	.678	-.13, .19
Segment 4	-.11	.254	-.30, .09	.00	1.00	-.16, .16
Segment 5	-.13	.186	-.31, .07	-.03	.744	-.18, .13
Derek						
Segment 1	.01	.882	-.18, .21	.02	.754	-.13, .18
Segment 2	.00	.995	-.19, .19	.04	.632	-.12, .20
Segment 3	.07	.493	-.13, .26	-.06	.422	-.22, .10
Segment 4	-.11	.253	-.30, .08	.04	.589	-.12, .20

Further point biserial correlations were run between the potential divergent validity marker, interpersonal trust, and the SAP (see Table 30). Again, the relationships ranged from

negligible to very weak, with only 5 of the 50 coefficients significant. Less trust was associated with increased likelihood of suspicion for Aaron segment 1, r_{pb} = -.14, p =.044, and Amy segments 1, r_{pb} = -.15, p =.037 and 3, r_{pb} = -.15, p =.035. Higher trust was associated with more suspicion for Jia segments 3, r_{pb} = .14, p =.047, and 4, r_{pb} = .17, p =.021.

Table 29

Point biserial correlation coefficients (Spearman's Rho) for the SAP and social vulnerability

	Behaviour			Suspicion		
Segment	*Rho*	*p*	95% CI	*Rho*	*p*	95% CI
Aaron						
Segment 1	-.03	.675	-.18, .12	-.05	.520	-.19, .10
Segment 2	-.06	.408	-.21, .09	-.03	.686	-.18, .12
Segment 3	-.06	.449	-.20, .09	-.05	.506	-.19, .10
Segment 4	.03	.665	-.12, .18	.00	.958	-.14, .15
Segment 5	-.05	.527	-.19, .10	-.12	.095	-.26, .03
Andy						
Segment 1	-.03	.671	-.18, .12	-.08	.255	-.23, .06
Segment 2	-.15	.035	-.29, -.01	-.13	.071	-.27, .02
Segment 3	-.07	.342	-.22, .08	-.02	.780	-.17, .13
Segment 4	-.16	.023	-.30, -.02	-.07	.336	-.21, .08
Segment 5	-.09	.202	-.24, .05	-.09	.210	-.23, .06
Segment 6	.00	.992	-.15, .14	.06	.383	-.08, .21
Amy						
Segment 1	-.13	.066	-.27, .01	-.06	.381	-.21, .08
Segment 2	-.02	.753	-.17, .12	-.16	.021	-.30, -.02
Segment 3	-.10	.161	-.24, .05	-.04	.583	-.18, .11
Segment 4	.04	.588	-.11, .19	.06	.439	-.09, .20
Segment 5	.00	.956	-.14, .15	-.11	.144	-.25, .04
Jia						
Segment 1	-.05	.538	-.19, .10	-.05	.454	-.20, .09
Segment 2	-.07	.319	-.22, .07	-.09	.196	-.23, .05
Segment 3	-.12	.111	-.26, .03	-.10	.174	-.24, .05
Segment 4	-.10	.175	-.24, .05	-.15	.040	-.29, .00
Segment 5	-.10	.222	-.25, .06	-.08	.250	-.23, .06
Derek						
Segment 1	-.15	.032	-.29, -.01	-.11	.142	-.25, .04
Segment 2	-.11	.131	-.25, .04	-.11	.113	-.26, .03
Segment 3	-.13	.073	-.27, .02	-.13	.073	-.27, .02
Segment 4	-.01	.857	-.16, .13	-.10	.156	-.24, .04

Table 30

Point biserial correlation coefficients (Pearson coefficient) for the SAP and trust

	Behaviour			Suspicion		
Segment	*r*	*p*	95% CI	*r*	*p*	95% CI
Aaron						
Segment 1	-.06	.402	-.20, .08	-.14	.044	-.28, .00
Segment 2	-.02	.788	-.16, .12	-.02	.767	-.16, .12
Segment 3	-.05	.517	-.19, .10	-.02	.812	-.16, .12
Segment 4	.06	.423	-.08, .20	.05	.447	-.09, .19
Segment 5	.03	.682	-.11, .17	-.07	.334	-.21, .07
Andy						
Segment 1	-.05	.499	-.19, .09	-.04	.549	-.19, .10
Segment 2	.07	.364	-.08, .21	.14	.052	.00, .28
Segment 3	.06	.458	-.09, .20	.01	.886	-.13, .15
Segment 4	-.03	.647	-.17, .11	.00	.984	-.14, .14
Segment 5	.13	.069	-.01, .27	.11	.133	-.03, .24
Segment 6	-.07	.345	-.21, .07	-.01	.916	-.15, .13
Amy						
Segment 1	-.12	.110	-.25, .03	-.15	.037	-.28, -.01
Segment 2	-.08	.264	-.22, .06	-.05	.481	-.19, .09
Segment 3	.10	.182	-.05, .24	-.15	.035	-.29, -.01
Segment 4	.01	.921	-.14, .15	-.03	.664	-.17, .11
Segment 5	.06	.408	-.08, .20	-.10	.175	-.23, .04
Jia						
Segment 1	.01	.945	-.14, .15	.00	.947	-.15, .14
Segment 2	.08	.298	-.07, .22	.00	.946	-.15, .14
Segment 3	.09	.221	-.05, .23	.14	.047	.00, .28
Segment 4	.08	.280	-.07, .22	.17	.021	.03, .30
Segment 5	.05	.557	-.11, .20	.08	.260	-.06, .22
Derek						
Segment 1	-.01	.843	-.15, .13	-.07	.345	-.21, .07
Segment 2	-.03	.728	-.17, .12	.06	.440	-.09, .20
Segment 3	.04	.567	-.10, .18	-.02	.783	-.16, .12
Segment 4	-.10	.181	-.24, .05	.03	.695	-.11, .17

Discussion

Studies 1 and 2 demonstrated that the SAP was an effective way of collecting detailed information about how people detect and respond to suspicious behaviour within several social interactions, and that consistent response patterns were elicited across multiple samples from the general population. The present study contributed several key additional findings.

First, SAP responding was remarkably similar for autistic and non-autistic participants. Within most of the crime scenarios, the proportion of participants reporting adaptive behaviours and suspicion responses in both groups increased as the scenarios progressed, whereas in the non-crime scenario suspicion remained low while adaptive behaviours varied across the scenario.

Second, few significant relationships were detected between ToM and both adaptive behaviours/suspicion. There was, however, one critical factor that likely mitigated against detecting such relationships. Although, at the group level, the autistic participants performed significantly worse on the measure of ToM, a substantial proportion of them performed at or near ceiling level—as was also the case for the non-autistic participants—while only a relatively small proportion received low scores. However, when participants at the extreme ends of ToM, but with similar levels of verbal ability, were compared, those with very low ToM scores were less likely to respond adaptively and report suspicion either as early in the different scenarios or as often as those with relatively high ToM scores. In other words, these post hoc examinations of extreme scorers suggest that pronounced ToM difficulties—irrespective of the presence or absence of an autism diagnosis— were associated with difficulties in discriminating suspicious behaviour and responding adaptively to such behaviour.

Third, in several scenarios a perhaps surprising pattern emerged with higher autistic traits associated with a greater likelihood of adaptive responses and increased suspicion. Fourth, several relationships emerged between the SAP and verbal ability, with participants with higher verbal ability generally more likely to respond adaptively or report suspicion.

Finally, relationships between the SAP performance indices and each of the potential concurrent and divergent validity measures were mainly negligible or, at best, extremely weak, with the few significant relationships quite likely being chance occurrences

considering the large number of relationships examined. In the following sections I examine each of these main findings in more detail.

The Relationship between Autism Diagnosis and SAP Performance

Although there is inadequate empirical evidence to support the argument that impaired ToM is a risk factor for criminal vulnerability specific to autistic defendants, it is frequently raised as a mitigating factor in defence arguments and expert testimony. Consistent with the findings of Brewer et al. (2023), the present research did not find evidence to support the argument for an autism-specific difficulty in discriminating suspicious or dodgy behaviour. Instead, autistic and non-autistic participants performed in a remarkably similar manner on the SAP, with differences in responding between the groups detected on only four occasions (Andy segment 2 suspicion, Amy segment 3 suspicion, Jia segment 2 behaviour and Derek segment 4 suspicion). Notably, on each of these occasions, the autistic participants reported more adaptive behaviours and suspicion than the non-autistic participants, although the differences were minor.

There are several possible explanations for the finding of no group difference. First, the format of the task and its repeated-measures design may have facilitated the detection of patterns within the scenarios that would be much harder to discern in spontaneous and complex real-world interactions, with the patterns identifiable by autistic and non-autistic participants alike. Previous research has suggested that autistic individuals can deduce the answers on lab-based tasks via analytical reasoning (Frith, 1994). Therefore, performance in such studies may not reflect the difficulties individuals may experience in real-world interactions. Despite its apparent greater ecological validity than the dodginess discrimination paradigm developed by Brewer et al. (2023), the SAP may not simulate many of the potentially critical but subtle social cues embedded in real-life interactions and, therefore, may not have exceeded the ability of the participants involved. For example, four autistic

participants mentioned a ‘freeze response’ at different points throughout the task, indicating they would not know how to respond at that moment (e.g., “I would be scared that I have been set up and I am in trouble and freeze paralysed in fear”). While this did not have significant adverse implications within the study, such reactions could result in much poorer outcomes in real life.

Alternatively, relevant prior life experiences and social skills training may have shaped the participants’ responses. For example, one autistic participant messaged the research team after participating to provide further information regarding their vulnerability. They wrote, “[I have a] whole developed moral code built up over the years, that I’ve had to train myself into following, and I’m trying to now consider how I would have reacted as my child/younger self to the same scenarios. I have sadly been in 3 out of those 4 scenarios and have gotten myself into trouble with my past naivety, trying to do the right thing. I struggle with recognising danger [and] facial expressions”. Participants may have developed compensatory strategies to manage their difficulties with social cognition earlier in life and applied these to the present task. For example, within the Aaron scenario, at the point where Aaron’s friend insists that the pictures of young children are for a photography class but that it is a surprise for his wife, so she cannot know, one autistic participant responded, “this would concern me, I would contact a friend for advice” with the explanation: “one of my coping strategies in life is to develop close friendships and contacts across a wide range of sectors, including the police, medical, mental health, social work. The reason I have done this is because of scenarios such as the one being commented on. So right at this moment I would be asking my friend at the police to help me go to the police, this serves 2 purposes 1) I need help in describing stuff in these situations so my friend is a hand-holder, 2) if I’m unsure of the moral framework, I can rely on the wisdom of another.” Such responses suggest that certain prior experiences and the development of coping strategies through social learning may provide a

protective factor for autistic adults in the situations simulated in the SAP. Of course, this is likely the case for all individuals—whether autistic or non-autistic— who participated in the study as social learning is a universal experience. However, specific interventions, memorable life experiences or compensatory strategies, such as those mentioned above, may provide some indication as to why no group differences emerged despite the literature documenting increased social vulnerability and victimisation of autistic individuals compared with non-autistic adults (Griffiths et al., 2019; Maïano et al., 2016).

Some experimental evidence for social learning in autistic adults was detected when Van Tiel et al. (2021) investigated strategic deception in adults with autism. They used a computerised game to examine whether autistic and non-autistic adults differed in their use, and detection, of deception against a computerised opponent. Participants were required to move their marker (a red circle) around a 5 x 5 grid, with the goal of capturing treasure (a grey square). The player was also instructed to capture the treasure before the computerised opponent (a green circle), who could not see the treasure. The player could use deception to do so by first moving away from the treasure to deceive the opponent about its whereabouts. Points were awarded for the distance between the two players when the treasure was captured. There was both a passive phase and an active phase, which determined whether the participant or opponent could view the location of the treasure. When in the passive phase, the computerised opponent was programmed to deceive the player in the same way each time. Van Tiel et al. (2021) anticipated that difficulties in perspective-taking would be associated with impairments in the ability to deceive or detect deception. However, they also hypothesised that social learning could play a role in influencing autistic participants' responses as the game progressed. Specifically, they believed a learning effect would reveal that autistic participants' use of deception and deception detection would improve over time. They found evidence to partially support this prediction, with findings suggesting that autistic

participants were initially less likely to deceive but became equally likely to deceive toward the end of the game. It also took them a longer time to realise that they were being deceived than non-autistic participants. However, the predictions were only partially supported as the learning effect was only observed for strategic deception (not deception detection) and the increased response times were only observed for deception detection. Nevertheless, the results suggest that autistic participants may engage in compensatory social learning strategies to mitigate ToM difficulties.

In a more realistic setting, this perspective is further supported by research suggesting that social skills training programs improve overall social skills, social responsiveness, and social skills knowledge in autistic adolescents with long-term sustained benefits (Mandelberg et al., 2014). For example, the UCLA PEERS Program evaluated by Mandelberg et al. (2014) included didactic lessons targeting online safety, an area targeted by the Jia crime scenario in the current study. While many studies and meta-analyses have focused on the impact of early intervention in childhood and adolescence (Fuller & Kaiser, 2020; Soares et al., 2021), more recent evidence supporting social skills interventions for autistic adults has also been reported (Dubreucq et al., 2022). Considering the increase in awareness and resources dedicated to improving autistic quality of life over recent years, prior experiences and learned information may mitigate potential vulnerabilities, thereby reducing any group differences. However, although several participants provided detailed responses about their personal experiences, without access to information regarding all participants' prior life experiences and therapeutic interventions this is purely speculative. Collecting such information could be an avenue for future investigation that will be discussed in the General Discussion chapter.

An additional potential indicator of increased autism awareness was suggested by the average age at diagnosis reported by the autistic sample. Age at diagnosis ranged from two to 59 with a mean of 22, which indicated a large proportion of participants who received their

diagnosis beyond childhood/adolescence. It is possible that people diagnosed earlier or later in life could differ in their responding; however, post-hoc analyses demonstrated no consistent difference in SAP responding between autistic participants below the median age at diagnosis (<24 years) and those at or above the median age (≥24 years). When, if ever, age of diagnosis might become important in this context will require longitudinal investigation.

Of course, it is entirely possible that, at a group level, autistic individuals are no more vulnerable than non-autistic individuals to manipulation in the situations depicted. The current findings correspond with those of Brewer et al. (2023), who found no difference between autistic and non-autistic participants on signal detection theory measures of discrimination of "dodginess". This would seem a likely explanation due to the significant heterogeneity of the condition and, therefore, varying risk and protective factors for each individual. For example, while some participants may display particularly low ToM skills, the current study and prior research (Brewer et al., 2017; Brewer et al., 2022) have demonstrated that others clearly do not. Further implications of the non-pervasiveness of ToM difficulties among autistic adults will also be considered in the General Discussion.

Although the results primarily demonstrated that there were no group differences, they also revealed that, if anything, autistic participants were slightly more likely to report suspicion and respond adaptively. Perhaps they were slightly more sensitive to a perceived demand to report suspicion, which is reflected in a group difference at the final segment of the non-crime scenario. Similarly, if the autistic participants believed themselves to be more gullible, as the group contrasts on the gullibility measure indicated, it is possible that they displayed extra vigilance during the task. Although these are possibilities, it is worth noting that Brewer et al.'s (2023) signal detection approach detected no group difference in response bias, that is, the tendency to report dodginess. It is also important to note that group

differences emerged in only four out of 50 potential responses with small effect sizes, so any differences were minimal.

The Relationship between SAP Performance and Autistic Characteristics

In Study 2, three relationships emerged between autistic characteristics in the since-removed Casey scenario, while no relationships emerged in other scenarios. The present study revealed relationships between autistic characteristics and SAP performance in the Aaron, Andy, Jia, and Derek scenarios. Contrary to expectation, in all cases across both studies, higher autistic traits were associated with more adaptive behaviours and suspicion. This finding is curious considering items on the AQ-12 specifically target social interaction rather than other facets of autism (e.g., restricted and intense interests), and does not align with (a) Williams et al. (2018) finding that autistic characteristics were negatively associated with performance on a realistic lie detection paradigm, or (b) Brewer et al.'s (2023) finding that autistic traits were unrelated to the discrimination of dodginess.

It is possible that explanations similar to those for the few autistic-nonautistic group differences on the SAP are relevant in this case. For example, further to the previously advanced notion that autistic participants high in self-reported gullibility may have been extra vigilant in their responding, so too might be the case for those high in autistic characteristics. In other words, self-awareness of one's social difficulties may have translated to diligence when completing the SAP. At odds with this explanation are the findings of prior research that autistic individuals are not more conscientious (Lodi-Smith et al., 2019) than non-autistic individuals. Thus, further research will be needed to address this issue.

The Relationship between ToM and SAP Performance

Further to the finding that SAP response patterns were so similar for both groups, the current study provided additional evidence consistent with Brewer et al.'s (2023) findings that ToM difficulties were associated with poorer discrimination and response to suspicious

cues, regardless of diagnosis. As noted previously, the likelihood of discovering meaningful relationships between ToM and SAP performance was constrained by the restricted range of ToM scores. Yet, several significant, albeit weak, relationships emerged in the Aaron and Andy scenarios. The placement of the segments in which the relationship emerged (Aaron segments 2 and 3; Andy segments 2, 3, 5, and 6), and the fact that the proportion of adaptive behaviours or suspicion at those segments was lower than other segments, suggests that ToM ability may be particularly important for detecting more subtle cues. For example, in segments two and three of the Aaron scenario, the protagonist is (a) asked whether he would like to store some files for a friend in exchange for help paying bills and (b) asked to keep those files private. It is perhaps unsurprising that participants with enhanced ToM were able to foresee potential risks associated with these statements. In the subsequent segment, where images of young children were mentioned, the cue was apparent enough to influence many participants' responses. This was shown by the high proportion of adaptive behaviours and suspicion in that segment (94.9% and 95.2%, respectively). Clearly, limited variability in both ToM and SAP responding meant that the more obviously suspicious segments were unlikely to reveal relationships. Similarly, in the Andy scenario, although relationships emerged at the final segments, the proportion of adaptive and suspicious responses remained low relative to the other crime scenarios. For example, in segment 5, 56.1% of participants responded adaptively, and 60.1% reported suspicion. In fact, across the entire scenario, the proportion of either behavioural category never exceeded 63.9%. This potentially allowed for additional ToM-SAP relationships to emerge due to more even distribution of responses between the adaptive/maladaptive and suspicious/not suspicious categories.

The current study also provided evidence of a link between extreme ToM scores and the detection of suspicious cues. Specifically, those with more marked ToM difficulties were less likely to respond adaptively as often and reported suspicion at later stages in scenarios

than those with advanced ToM, although the two sub-groups were of similar verbal ability. When examining the change across segments when the crime scenarios were combined, differences between the 'high' and 'low' ToM sub-groups emerged. For example, in the first segment, many participants in both the low and high ToM groups reported two or fewer adaptive (86.4% and 76.8%, respectively) or suspicious responses (92.3% and 86.7%, respectively). However, as the segments progressed, the proportion of high ToM participants providing more than two adaptive responses increased from 23.2% at segment 1 to 78.4% at segment 5. In comparison, the low ToM group only increased from 13.6% to 45%. A similar pattern was observed for suspicion responses, where the high ToM group moved from 13.3% to 90.2% while the low ToM group moved from 7.7% to 67.9%. In other words, as the scenarios progressed, the participants with high ToM were more likely to report adaptive behaviours or suspicion than those with very low ToM scores.

The Relationship Between Verbal Ability and SAP Performance

Unlike the A-ToM-Q data, there was no restriction of range on the proxy VCI measure. The current data revealed several relationships between SAP responding and VCI, with participants with higher verbal ability more likely to respond adaptively and report suspicion[6]. This outcome is consistent with Brewer et al.'s (2023) finding that verbal ability predicted dodginess discrimination, independent of ToM. It is unsurprising that verbal ability was associated with SAP responding, considering the paradigm is an entirely audio-based format which (a) prevented participants from exploiting visual cues such as those that might be present in many real social interactions, and (b) required adequate language comprehension to detect any cues to suspicious behaviour.

[6] One exception to this trend emerged at the beginning of the Aaron scenario, where higher verbal ability was associated with fewer extricating behaviours (rpb = -.16, p =.029).

The Relationship Between the SAP and Potential Markers of Concurrent and Divergent Validity

The relationships between SAP performance and the putative validity markers failed to provide independent confirmation of the SAP's convergent validity. All correlation coefficients were either negligible in size or extremely weak. Although some were significant, albeit weak, this was not unexpected given the number of correlations reported.

Considering these results, what should be inferred regarding the validity of the SAP? One possibility is that the SAP does not provide a genuine index of the ability to detect suspicious or dodgy behaviour. An alternative explanation is that the validity markers I selected are inadequate. There are several reasons for believing the latter explanation may be accurate. First, the measures are self-report and may not be veridical indices of likely behaviour. Second, the validity data for the gullibility scale are, with one exception, based on other self-report measures (e.g., Paranormal Beliefs Scale; Tobacyk, 2004), Social Intelligence Scale; Grieve & Mahar, 2013). Third, on closer examination when trying to account for the absence of relationships, it became apparent that the only criterion-related behavioural validity data reported for the gullibility scale are problematic.

George et al. (2020) reported that individuals who scored highly on the gullibility scale were more likely to click on simulated phishing emails (e.g., verification or account issue notifications from companies such as Netflix, Commonwealth Bank, Facebook, and the Australian Tax Office) and provide their details. Closer examination reveals, however, that this conclusion was based on a highly selective parsing of the data. The authors reported comparisons of gullibility scores for responders (i.e., people who engaged with the phishing emails) and non-responders (i.e., people who did not engage with the phishing emails) using three different samples or subsets of their respondents. The first comparison used the total sample (37 responders, 182 non-responders) returned a non-significant difference in

gullibility scale scores between those engaged with the emails and those who did not. Surprisingly, the authors disregarded this finding on the grounds that participants who may have deleted the simulated phishing email before opening it would have been included in the non-responders category (George et al., 2020, p. 3). Given the exhortations we all receive not to open emails we are unsure about, this "massaging' of the data seems inappropriate. Next, they compared responders (n = 17) and non-responders (n = 88) to a follow-up survey and again found no significant difference in gullibility. Again, they ignored this null finding on the grounds—albeit with no evidence—that some responders might not have remembered receiving the emails. In their third comparison, the authors compared the gullibility of responders (n = 6) and non-responders (n = 34) who confirmed receiving the four phishing emails in the follow-up survey. This (now) significant comparison was deemed by the authors as the most reliable and the basis for claiming criterion-related validity for the measure. In other words, it is reasonable to argue that George et al. (2020) does not provide the evidence of criterion-related validity that they claimed.

An additional explanation for the lack of relationships between the SAP and the gullibility measure may lie in the specificity of gullibility as described by (Teunisse et al., 2020), where they contend that gullibility is conceptually distinct from compliance in that "an individual may or may not act upon this acceptance, but it is the acceptance of the false premise despite the presence of untrustworthiness cues that is central to the concept of gullibility" (p. 409). Within the SAP, in addition to poor suspicion recognition, participants could demonstrate compliance by reporting maladaptive behaviours (e.g., accepting an offer or complying with a request) despite recognising the potential for suspicious activity. For example, in the Aaron scenario, a participant indicated, "I would choose to hold the pictures even though this seems off" with the explanation, "because I need the money, and he's offering to pay my bills". Furthermore, responses could also reflect participants' rigidity in

specific scenarios, where they perhaps failed to recognise suspicion but responded adaptively for other reasons (e.g., “I would not take the job” with the explanation “because I do not like driving”). Therefore, the conceptual association I initially predicted between the SAP and gullibility may have been partially erroneous, as the behavioural responses on the SAP reflect additional decision-making processes such as compliance or rigidity.

A crude investigation of compliance was undertaken by analysing the frequency of misalignment between behaviour and suspicion responses (i.e., when participants indicated that they were suspicious but reported a maladaptive behaviour, that is they ‘complied’ with the person’s request) for those participants where no SAP data were excluded (n = 108). This revealed no group differences in the total number of ‘compliant’ responses across the 25 segments (autistic: M = 2.00, SD = 1.73, Mdn = 2.00; non-autistic: M = 1.91, SD = 1.61, Mdn = 1.00), t (106) = -.28, p=.778, d = -0.05, 95% CI [-0.43, 0.32], U = 1054.50, p = .767. Nevertheless, considering this was not a comprehensive analysis, future studies could further investigate whether gullibility, compliance, rigidity, or other cognitive processes were particularly influential in driving participants’ decision-making. These and other potential future directions will be further discussed in the General Discussion.

The emergence of three relationships between interpersonal trust and the SAP (Aaron and Amy segments 1; Amy segment 3), although unexpected, were more easily understood. Teunisse et al. (2020), along with previous researchers (Rotter, 1967; Yamagishi et al., 1999), contended that interpersonal trust exists independently from gullibility, as gullibility requires the presence of untrustworthy cues while trust requires only a dispositional belief that others’ word can be relied upon. As the first segment of each scenario in the SAP was designed to be relatively neutral or unsuspicious, it is conceivable that participants low in general interpersonal trust would be more suspicious of those interactions. For example, in segment 1 of the Amy scenario, while making small talk at her checkout, the protagonist is asked by an

older male customer whether she attended a local school as she looks familiar. In the Aaron scenario, the protagonist is told by his boss that his shifts at work have been cut due to the pandemic. Individuals low in trust may be more likely to interpret these cues, without additional context, as being suspicious due to a general distrust of others. It was then surprising that higher trust was associated with increased suspicion at Jia segments 4 and 5; however, it must also be acknowledged that the likelihood of some relationships emerging due to chance was reasonably high given the 25 individual segments (and 50 responses) that comprise the SAP.

Taken altogether, queries such as those I have raised in this section about the relationship between the SAP and the potential validity markers examined reflect the need for obtaining objective 'real-world' criterion-related validity data. In the next chapter I discuss possible limitations on generalising from the SAP data to discrimination of, and adaptive responding to, suspicious behaviour in everyday settings.

Conclusions

The findings of Study 3 did not indicate that autistic adults were less likely than non-autistic adults of similar verbal ability to become suspicious or respond adaptively than non-autistic individuals in the scenarios presented within the SAP. Furthermore, there was no indication that they needed more information to do so: that is, they were no more likely to pick up on the suspicious cues at a later stage than the non-autistic participants. The study also supported the argument that ToM difficulties are not a ubiquitous feature of autism in adults and that, regardless of diagnosis, very poor ToM may render some individuals less likely to become suspicious and respond adaptively. Overall, the findings were largely consistent with those reported using quite a different paradigm in the Brewer et al. (2023) signal detection study.

CHAPTER 5

General Discussion

Despite the absence of compelling empirical evidence supporting the argument that impaired Theory of Mind is a risk factor for criminal vulnerability specific to autistic defendants, it is frequently raised as a mitigating factor in defence arguments and expert testimony. For example, in the case of R v Middleton (2023), a 22-year-old male was arrested for possessing and distributing child exploitation material. In his defence, the argument was made, and accepted by the court, that his autism made him "vulnerable to meeting people online who could take advantage of his underdeveloped social and emotional maturity" and that "his naivety and literal interpretations meant that if a user profile stated that the person was a 25-year-old female, he genuinely thought he was talking to someone of that age and gender" (para. 56). Although anecdotal evidence certainly suggests there may be characteristics of autism that contribute to criminal vulnerability and victimisation (Brewer & Young, 2015; Freckelton, 2013; Freckelton & List, 2009; *R v Middleton,* 2023), there is limited empirical evidence regarding the social-cognitive processes that may underlie such vulnerabilities. The present research investigated whether ToM—hypothesised to be a core feature of autism— undermines the ability to detect suspicious behaviour within an interaction and how that might impact vulnerability to naïve criminal involvement.

Building upon previous research in this area (Brewer et al., 2023; Brewer et al., 2018; Young & Brewer, 2020), I developed and evaluated a novel paradigm—the SAP—designed to measure people's ability to detect and respond to suspicious behaviour within social interactions. The first phase of this project involved developing and refining the SAP with large samples of non-autistic adults to understand how participants would respond to several social scenarios. This involved (a) identifying examples and non-examples of a broad range of behaviours that could underpin a comprehensive coding protocol, (b) establishing the

reliability of the protocol via repeated inter-rater reliability analyses and discussions, and (c) showing relatively stable patterns of responding across participants that indicated the sensitivity of the paradigm to the proposed cues to suspicion. The project's second phase then included comparison measures for a preliminary exploration of the SAP's association with ToM, autistic traits, and verbal ability. This study generally found no relationships between ToM or autistic traits and the SAP but highlighted significant restrictions in variability on the ToM measure. Furthermore, it indicated that, as expected, verbal ability had a meaningful association with SAP performance. The third and final phase involved adding an autistic sample to investigate autistic-non-autistic group differences and further explore associations between the SAP and ToM, autistic traits, and verbal ability. It also included additional measures of gullibility, social vulnerability, and interpersonal trust as potential markers of concurrent and divergent validity.

The Association between Autism Diagnosis and the Detection of Suspicious Behaviour

Prior research exploring the connection between the ability of autistic and non-autistic individuals to detect suspicious behaviour has yet to yield convincing evidence of group-level differences. Williams et al. (2018) reported that autistic participants were significantly less accurate at detecting lies in a deception detection task when compared with an IQ-matched neurotypical sample. However, although similar, the relevant constructs—deception detection and the recognition of suspicious behaviour—are not synonymous. For instance, Williams et al.'s (2018) paradigm required the binary identification of truth-tellers or liars after watching video stimuli of people accused of cheating in an earlier research task. Although 20 stimuli were used, the videos displayed various people responding to accusations within a single context (i.e., all stimuli participants were accused of cheating within the same research task, using the same line of questioning). Therefore, it is possible that non-autistic raters were better able to recognise the relevant cues for that specific context. Alternatively, it may be

that a categorical judgement about whether somebody is lying or not involves quite different cognitive processes than predicting whether an interaction is becoming problematic. For example, deception detection research reports consistently poor performance by general population samples when distinguishing truths from lies (e.g., around 54% accuracy; Bond & DePaulo, 2008). Within the SAP, the vast majority of participants across several large samples identified suspicion and responded adaptively by the conclusion of each scenario, suggesting that the cognitive processes underlying each task are quite different.

Using a signal detection theory approach, Brewer et al. (2023) found equivalent discrimination of suspicious cues and no response bias (i.e., a tendency to just report suspicion) in both autistic and non-autistic participants. Study 3 of the current project found further evidence to support this finding using a more ecologically valid paradigm. It showed an apparent absence of group differences in adaptive behaviours and suspicion responses across crime and non-crime scenarios, as well as no tendency for either group to be more likely to report suspicion in a non-crime scenario. As discussed in Chapter 4, several potential explanations could be offered for these results.

First, perhaps no group-level vulnerability exists in autistic adults. This explanation is consistent with the literature regarding prevalence estimates of criminal vulnerability in autistic adults, a literature that suggests that autistic adults are no more likely to offend than non-autistic adults (Weiss & Fardella, 2018; Yu et al., 2021; though see Collins et al., 2023). Second, it is possible that participants' responses were shaped by social learning, relevant prior life experiences or social skills training. Third, the design of the SAP may not have captured the nuance of social interactions (e.g., body language, facial expressions) that would reveal vulnerabilities that would present in many real-life interactions. This idea is discussed further in the chapter's Limitations and Future Directions section.

Although the SAP generally demonstrated a lack of group differences, there were several occasions on which the autistic participants reported more adaptive behaviours and suspicion than the non-autistic participants. This finding corresponded with the discovery that higher autistic traits were sometimes associated with more adaptive behaviours and suspicion responses in Studies 2 and 3. As discussed in Chapter 4, these results are unexpected and at odds with previous research (Brewer et al., 2023; Williams et al., 2018) and highlight a need for further research in this area. Perhaps, in some cases, and particularly within this paradigm, autism can provide a protective factor from vulnerability. For instance, is possible that alternative autistic characteristics, such as insistence on sameness, rigid moral code, or limited social motivation (e.g., not wanting to drive people around because of an aversion to small-talk or spending time with strangers), could have prevented autistic participants from following a potentially perilous path (Howlin, 2004; Mouridsen, 2012). The reduced-item version of the measure of autistic traits used in Study 3 does not capture such characteristics in detail, and omits others altogether (e.g., intense interests or repetitive behaviours). However, considering a minor relationship in the same direction emerged in Study 2 using the AQ50 on a non-autistic sample (and therefore likely less variability than a group comparison), this relationship clearly warrants further investigation. The potential influence of additional characteristics of autism is further discussed in the Limitations and Future Directions section.

ToM and the Detection of Suspicious Behaviour

Beyond the exploration of group differences, the current project investigated the relationship between ToM and the ability to detect and respond to suspicious behaviour. It was anticipated that the addition of an autistic sample would allow for increased variability on the ToM measure compared with Study 2. However, Study 3 also found that that ToM difficulties were not pervasive among autistic participants, with many autistic participants

performing at or near ceiling—a finding that reflects concerns expressed by some researchers about designating ToM difficulties as a core feature of autism in adults (Astle & Fletcher-Watson, 2020; Brewer et al., 2017; Brewer et al., 2022; Gernsbacher & Yergeau, 2019). Thus, despite the inclusion of the autistic sample, the lack of variability on the ToM measure continued to constrain the ability to detect meaningful relationships. Nevertheless, several relationships emerged across Studies 2 and 3, indicating that difficulties with ToM were associated with fewer adaptive behaviours and suspicion. This association became more salient when the performance of low and high ToM sub-groups was contrasted. Those analyses revealed that, as the scenarios progressed, participants with very low ToM, irrespective of diagnosis, were less likely to report adaptive behaviours and suspicion at the same rate as those with very high ToM. This suggests that although ToM is potentially a risk factor for criminal vulnerability, those with the most significant ToM difficulties are likely to be most at risk. This finding echoed Brewer et al.'s (2023) signal detection study, in which ToM was independently associated with the discrimination of 'dodgy' or suspicious cues. Similarly, Young and Brewer (2020) found that ToM was related to the ability to provide extricating information to authorities when erroneously accused of a crime. Conversely, in their deception detection study, Williams et al. (2018) did not find evidence of a relationship between ToM (referred to as mindreading ability) and deception detection. However, ToM was not examined in their group contrast study where perhaps more variance on ToM might have been expected. Additionally, in their first study using a non-autistic sample, there may have been insufficient participants performing poorly on the ToM measure. The ToM measures used in their study (Reading the Mind in the Eyes task; Baron-Cohen et al., 2001, and the Animations task; Abell et al., 2000) have also demonstrated problematic discrimination between groups with adult samples (Brewer et al., 2017) and limited validity (Quesque & Rossetti, 2020).

Although the present study provides a promising indication that pronounced ToM difficulties, regardless of autism diagnosis, may be associated with the ability to detect suspicious behaviour and respond adaptively, it also must be acknowledged that the relationship between ToM and the detection of suspicious activity is undeniably complex. For example, ToM difficulties do not necessarily mean that an individual is incapable of detecting suspicious behaviour but rather that they may find it more challenging in certain situational conditions. Furthermore, detecting suspicious behaviour within social interactions likely involves a range of cognitive abilities and factors, such as executive functioning and awareness of social norms. It remains essential to consider the influence of situational context, social pressures, and other individual differences that were not measured in the current research. These factors will be further discussed in the Future Directions section.

Implications for Criminal Responsibility of Autistic Defendants

In this section I discuss the implications of the main findings from the current project when considering the possibility that the very nature of autism may render autistic individuals vulnerable to involvement in crime as has sometimes been argued in legal cases. Three findings are particularly relevant to these considerations. First, consistent with other recent findings (Astle & Fletcher-Watson, 2020; Brewer et al., 2017; Brewer et al., 2022; Gernsbacher & Yergeau, 2019), ToM difficulties did not characterise all autistic adults. This suggests that it is judicious to emphasise individual characteristics within the context of an autism diagnosis when considering cases involving autistic defendants (O'Sullivan, 2018). As stated by Wolf (2021), "It is crucial that courts are wary of assuming that the symptoms that health practitioners identify as fundamental to their diagnoses of ASD necessarily or directly correlate with a risk of criminal offending" (p. 1704). For example, while suggesting that a person's ToM difficulties contributed to their offending may be entirely appropriate in some cases, evidence of ToM difficulties in those individuals beyond the diagnostic label is

necessary. One such solution would be the objective measurement of ToM using validated psychometric measures; however, the limitations of current ToM measures merit a focus on the development of more sophisticated instruments that capture the intricacies of ToM or other social-cognitive difficulties that might create problems in social interactions. For example, it could be valuable to differentiate the cognitive and affective components of ToM in future measures (Karoğlu et al., 2021).

Second, and also consistent with other recent studies (Brewer et al., 2017; Brewer et al., 2022), the ToM data indicate that the prevalence of quite marked ToM difficulties was higher in autistic than non-autistic adults. This emphasises that the consideration and assessment of ToM as a risk factor for autistic defendants remains essential, despite the finding that ToM difficulties were not observed in all autistic participants. Third, as found by Brewer et al. (2023), the ability to detect and respond to suspicious behaviour within interactions did not distinguish autistic and non-autistic adult samples. Therefore, the idea that an autism diagnosis is associated with an inherent inability to read cues to suspicion, and respond adaptively, was not supported by these data. Fourth, there was an association between ToM difficulties and the detection of suspicion that saw participants with very low ToM, regardless of diagnosis, failing to pick up suspicious cues or respond adaptively early within the interactions. Thus, the data do show that people with significant ToM difficulties may experience problems in this area and consideration should be given to this when indicated.

In sum, these findings have indicated that a more nuanced understanding of ToM as a criminological risk factor for autistic adults is needed to reduce the likelihood of negative outcomes when autistic individuals interact with the different sectors of the justice system. For example, the perception that autism is a lifelong neurodevelopmental condition with pervasive ToM difficulties and cognitive rigidity could lead to overly harsh sentences for fear of recidivism. Conversely, beliefs that diagnostic criteria must be relevant to the same extent

for everyone with a diagnosis could inappropriately mitigate culpability for some individuals. It is essential that knowledge in this area continues to evolve to assist sentencing decisions and the development, or revision, of appropriate rehabilitation programs and supports.

Limitations and Future Directions

The SAP as a Measure of the Detection of Suspicious Behaviour

It remains important to highlight that the current data were correlational and cannot speak to any causal relationships between ToM (or other comparison measures) and the detection and response to suspicious behaviour. Furthermore, while these studies presented a novel opportunity to investigate a more ecologically valid method—compared with Brewer et al.'s (2023) signal detection paradigm—of measuring the discrimination of, and adaptive responding to, suspicious behaviour, I emphasise that there are several limitations on generalising the SAP data to decision-making in everyday settings.

As mentioned previously, the current studies did not provide convincing evidence of the convergent validity of the SAP using extant markers of gullibility, social vulnerability, and trust. Chapter 4 proposed some compelling explanations for these findings, including the reliance on self-report markers of validity and the problematic evidence used to assert the criterion-related validity of the gullibility measure. These limitations reinforce the critical need for independent behavioural or outcome markers that would inform criterion-related validation. Future research that addressed such issues successfully would render the SAP, or a refined version of it, an invaluable tool for investigating any fundamental limitations that may constrain individuals' ability to navigate potentially problematic social interactions.

Similarly, just as I have highlighted the limitations of self-report data in the potential validity measures examined, it is pertinent to note that the SAP is obviously not a "pure" measure of behaviour. Rather, it measures people's intentions to respond in a certain way. There is, of course, no way of knowing whether these responses would translate to similar

behaviours in real-life situations. Furthermore, the SAP, in its present form, cannot simulate the many complex interpersonal demands of natural social interactions and the influence of social pressures on decision-making. Due to the ethical and methodological difficulty in conducting simulated scenarios of this nature, developing appropriate behavioural measurements would likely be a complex and challenging undertaking—although perhaps unnecessary if future research with the current version of the SAP convincingly demonstrated impressive criterion-related validity relationships.

Nevertheless, several refinements of the design of the SAP could be considered in future research. First, previous research has suggested that people with ToM difficulties take longer to detect suspicious cues (Brewer et al., 2018) and that people with autism typically take longer to respond to ToM items, irrespective of accuracy (Brewer et al., 2022). Latency data were not collected for the present research as SAP responses were not time-restricted in order to allow for individual differences in typing speed when producing sometimes lengthy responses and comments. Further, due to the task length, participants may have taken short breaks throughout and, therefore, latency data could not be relied upon as an accurate marker of decision-making speed. It is possible that, while the autistic participants performed similarly on the SAP to non-autistic participants, the time taken to formulate their responses was longer. Delays in detecting and responding to suspicious cues would likely not translate to optimal performance in real social interactions. The design of a paradigm that incorporates realistic and time-sensitive responding could be another vital avenue for investigation.

A second refinement of the current SAP measure would be to incorporate visual cues rather than relying entirely on audio-based scenarios. This would improve ecological validity by allowing participants to incorporate subtle social cues readily available in real life but not in the SAP (e.g., facial expressions, body language, or behaviour). It is possible that the addition of subtle non-verbal visual cues would reveal group differences that did not emerge

on the SAP. This would mirror research suggesting that in order to reduce the disparity between performance in laboratory-based tasks and real-world difficulties, studies need to closely reflect the demands of real life (Frith, 2004; Frith et al., 2003; Ponnet et al., 2008; Roeyers et al., 2001). This could be partially achieved by translating the audio scenarios into video stimuli. A further step to achieving this goal would be via the use of virtual reality, which has recently demonstrated utility in improving social skills training for both individuals with autism (Ke et al., 2022; Kourtesis et al., 2023; Parsons & Mitchell, 2002; Yuan & Ip, 2018) and schizophrenia (Oliveira et al., 2021). However, further research in this area would necessitate consideration of the potentially confounding nature of increased reliance on executive functioning skills (e.g., mental flexibility and attention) rather than ToM ability alone. Nevertheless, as executive functioning skills play a vital role in how all people navigate life, and executive functioning difficulties are a common feature of autism (Demetriou et al., 2018; Ozonoff, 1995), the interaction between executive functioning and ToM would be an important consideration for future research.

A third potential development for the SAP would be the manipulation of factors that might increase our understanding of the impact of situational context. For example, the SAP depicted a mix of strangers, acquaintances, and close friends across the scenarios. An interesting and important consideration could be how the relationship between characters impacted responding. For example, in the final segment of the Andy scenario, where Andy was dropping a young male off in the city and was asked to wait around, many participants indicated that they would not leave the boy alone due to his age. Future research could observe any differences in decision-making if the character was depicted as an adult or youth and how the depiction of different characters impacts participants' responding. Alternatively, manipulation of the protagonist's life situation could be investigated. For example, in the Jia scenario where the protagonist was facing fear of deportation, manipulations could be

introduced to examine whether participants would be driven to apply to the online program if the character was a domestic student with alternative study options. Investigating such differences would allow teasing apart of the impact of ToM from that of the specific contextual cues presented by the scenarios.

Such research could tie into research investigating differences in moral reasoning between autistic and non-autistic adults (Bellesi et al., 2018; Grant et al., 2018; Spenser et al., 2015). It has been argued that moral reasoning, the ability to analyse and judge whether something is 'right' or 'wrong', develops and presents differently for autistic and non-autistic adults. Specifically, some research suggests that there is little to no group difference in the judgement of simple moral dilemmas, but when considering complex scenarios, autistic individuals demonstrate use of concrete and inflexible categories of right and wrong according to learned social rules (Buon et al., 2013; Shulman et al., 2012; Takeda et al., 2007). In Study 3, this potentially relates to the lack of vulnerability in the autistic sample, where perhaps the autistic sample paid more attention to the veracity of the request (i.e., in the case of Andy, being asked to drive strangers around for money) with less consideration for the context of the protagonist (i.e., desperate for money and unable to find other work). There are mixed opinions in the literature about the interaction between moral reasoning and ToM (Baird & Astington, 2004; Leslie et al., 2006) and, thus, this could be investigated in future research by also including questions on moral reasoning (e.g., intentionality, causality, responsibility etc.) at the conclusion of the scenarios.

Along similar lines, the SAP only depicted a brief selection of social situations (four crimes, one non-crime). It would be advantageous to broaden this scope in future research by including additional or alternative scenarios. This would allow investigation of vulnerability to certain types of criminal activity (e.g., cyber, financial, possession-related). For example, using a between-subjects design in which participants are randomly allocated to one of

several scenarios could allow for the exploration of vulnerability to particular crime types without the potential influence of priming from repeated measures. Some research suggests that autistic adults may be particularly vulnerable to crimes such as arson (Allely, 2019; Cashin & Newman, 2009; Mouridsen, 2012) and sexual offences (Kumagami, 2006). Although these may be difficult to replicate in an experimental study, such findings underscore the need for further exploration of the impact of context. Alternatively, the final segment in which participants were informed of involvement by authorities could be removed. This would increase the moral ambiguity of the situations to reduce priming across scenarios.

Historically, between-subjects studies with multiple stimuli have been challenging to achieve in autism research where limited access to samples of autistic participants constrains statistical power. However, the introduction of crowdsourcing platforms such as Prolific and the growing number of adults with autism who use the platform (2,586 active[7] participants as of September 2023) has allowed for more appropriately powered samples than has historically been the case.

Sampling Restrictions

Although Prolific has many benefits, it is also necessary to note the reliance on self-reported diagnosis in Study 3 and the inability to verify these claims independently. Nevertheless, pre-existing demographic information collected by Prolific (i.e., disclosure of an autism diagnosis) allowed for instant screening of participants whose study responses to questions about diagnosis did not match their existing data. That is, if a participant was directed to the survey targeting autistic participants because they had recorded a diagnosis on Prolific, but then responded to a survey question denying an autism diagnosis (or vice versa for the non-autistic survey), they were screened from participating. Details regarding the

[7] Prolific defines an active user as somebody who has used the platform within 90 days.

diagnosis (e.g., age at diagnosis, practitioner providing diagnosis) were also collected. Furthermore, discrimination between the autistic and non-autistic samples was supported by the large and significant group difference in scores on the measure of autistic traits. Considering the increased access to large samples of autistic participants that is offered using Prolific, the ability to confirm diagnoses formally was thought to be an acceptable trade-off.

Verbal Ability and the SAP

Another finding to emerge from the current research project was the relationship between the SAP and verbal ability. This association was anticipated because of the reliance on spoken dialogue and cues within the SAP. Further, previous research has documented it as an independent predictor of 'dodginess' discrimination (Brewer et al., 2023) and extrication ability (Young & Brewer, 2020).

The strength of the association between verbal ability and SAP performance speaks to the potential importance of intellectual ability in negotiating potentially problematic social interactions. Although I targeted individuals with verbal IQs of 85 and above, impaired intellectual functioning often co-occurs with an autism diagnosis (Matson & Shoemaker, 2009), with some estimates suggesting approximately 30% co-occurrence (Baio et al., 2018). Consequently, the current findings cannot be generalised to those autistic adults with verbal IQs below 85. It is, however, possible that the impact of ToM difficulties on the detection of suspicious behaviour may be more pronounced in individuals with relatively low verbal ability, which would correspond with literature on the victimisation of individuals with intellectual disabilities. When Wilson and Brewer (1992) compared victimisation rates of a sample of 174 adults with intellectual disability with population statistics from the region (collected from the Australian Bureau of Statistics), they found that the individuals with an intellectual disability were twice as likely to experience personal offences and 1.5 times more likely to experience household victimisation. Furthermore, victimisation for certain crimes

(i.e., assault, robbery, and sexual assault) was significantly higher for the sample with an intellectual disability than the population statistics. They also reported that victimisation varied with level of disability, with mild to moderate disability associated with increased susceptibility to both personal and property crime, while more severe disability corresponded with high personal victimisation but lower property victimisation. Fisher et al. (2016) later conducted a systematic review investigating the reported prevalence and risk factors for victimisation since the original Wilson and Brewer (1992) publication. They found reports of widespread susceptibility to a range of negative experiences (sexual abuse, assaults, intimidation, property destruction, vandalism) with multiple victimisations being a common occurrence. They also reported that individual risk factors such as poor decision-making skills, limited education, and appearance of vulnerability increased the likelihood of victimisation. Both studies suggest that it would be advantageous for future research to consider the co-occurrence of autism and intellectual disability when investigating vulnerability.

The Influence of Other Characteristics of Autism

The current research project focussed on the relationship between ToM and the detection of, and response to, suspicious behaviour. However, several other common characteristics of autism could potentially play a significant role in elevating criminal vulnerability and victimisation. For example, while ToM difficulties in isolation appear to increase risk to some degree, ToM difficulties, in combination with other autistic characteristics, may further increase risk.

Consider the hypothetical example of a young male who has a keen interest in vintage cars. A group of acquaintances he previously attended school with know of this keen interest. One night, after stealing a vintage car from a nearby neighbourhood, they drive it to his house and ask him to store it for them. After seeing the vehicle and failing to anticipate or interpret

the motive behind the offer, he accepts and is subsequently arrested. Perhaps an individual with ToM difficulties but limited interest in cars would turn down the offer. Conversely, a person with a keen interest in cars but a well-developed ToM would decline the offer out of suspicion. In this case, the combination of ToM difficulties and interest in the vehicle led to his ultimate imprudent decision to accept the offer.

Another characteristic that may impact criminal vulnerability in conjunction with ToM difficulties is the presence of hyper-or hypo-sensitivities (Kellaher, 2015). For example, Brewer and Young (2015, p. 129) detailed a case in which a young autistic man named Bradley was charged with attempted kidnapping. Bradley had become agitated by his bus schedule changing and walked to a nearby primary school to ask for directions home. Two young girls were playing near the primary school, and Bradley approached them to display his rock collection. When the two girls screamed, Bradley reached out intending to quieten them—he had no intention of causing harm—and caused the girls even more distress. Bradley's poor ToM hindered him from anticipating the girls' reaction to him, and his hypersensitivity to noise subsequently influenced his response to try and quiet them, further aggravating the situation and resulting in criminal charges of attempted kidnapping.

A third possibility is that autistic individuals who are seeking to achieve a relationship with someone might be so focused on fostering this connection that their ToM difficulties become more evident. As discussed in Chapter 1, many autistic individuals report a desire for friendships and intimate relationships but have difficulty achieving and maintaining these (Mazurek, 2014; Orsmond et al., 2004; Strunz et al. 2017). Social isolation coupled with a desire for connectedness may further accentuate ToM difficulties by clouding cues to suspicion or malice with hope or optimism. In the case of R v Middleton (2023), it was reported that the defendant described himself as "socially isolated" and would "typically only engage in casual interactions with acquaintances via online platforms" (para. 43).

Furthermore, his psychological assessment noted that his "late awareness of a deficit in life experiences relating to romance or sex compared to his same-age peers made him feel defective and socially isolated, which was a precipitating factor in the commission of the offences" (para. 56). The defendant had noted that he had no sexual attraction to children, but that he opened many of the videos to confirm that they were child abuse material so that he could swap them with other people online, as "that is how [he] made friends" (para. 58).

Finally, a fourth characteristic of autism that was hinted at in the current findings was the possibility that a tendency towards compliance or acquiescence might contribute to vulnerability. This tendency was manifested on the SAP in responses that demonstrated maladaptive behaviours despite the recognition of suspicious behaviour. For example, responses such as "I would continue to store the files", with the explanation "it sounds a bit dodgy, but he is my friend, so I would help him out", indicated that, on rare occasions, participants were able to detect that the situation was problematic but acquiesced regardless. Although this pattern did not appear pervasive in the current research, it has potential ramifications for criminal vulnerability in real life, especially if the individual has difficulty reading the intentions of interaction partners. For example, another case presented by Brewer & Young (2015, p. 122) detailed the story of a young man named 'Garrison' who was arrested and charged with stealing copper piping from new housing developments. The offending occurred over several months together with a co-accused. Due to family difficulties, Garrison's co-accused had been housing him after he left home. In return, Garrison was asked to pay his entire government allowance to the co-accused. Garrison admitted under questioning that he had found it "a bit weird" that his co-accused had been carrying bolt cutters and a bag on their drives but did not think to question the co-accused when he began stealing copper piping from the properties. Garrison was then asked to begin stealing the piping himself. He then reportedly recognised that the activity was wrong but

complied with his co-accused's repeated requests and was unable to extricate himself from the situation. Some studies have shown that autistic adults may be more likely to comply to avoid conflict and confrontation within interactions (Chandler et al., 2019; North et al., 2008), while others have found no group differences (Maras & Bowler, 2012). The few studies in this area have been limited by small samples or reliance on self-report measures. Therefore, additional research is necessary to understand the relationship further. Including a self-report measure of compliance, such as the Gudjonsson Compliance Scale (Gudjonsson, 1989) or a behavioural measure (e.g., door-in-the-face-technique; Cialdini et al., 1975), in future research could be an effective way to begin exploring possible interactions between SAP performance and the tendency towards compliant responding. Importantly, although individual case studies suggest that the possibilities mentioned above are all plausible, they clearly require empirical investigation.

Conclusions

Previous research (Brewer et al., 2023; Brewer et al., 2018) has shown preliminary evidence of a relationship between ToM and the detection of suspicious behaviour but no evidence of group differences when comparing autistic and non-autistic adults. The current research expanded upon this work by developing a novel paradigm, the SAP, and subsequently examining the relationship between autism, ToM, and the ability to detect and respond to suspicious behaviour within interactions. The findings generally replicated previous findings, suggesting that autistic adults may not be inherently less likely to detect deception or "dodginess" than non-autistic individuals. This result has significant implications for the legal context, as it raises questions about the viability of using a generalised 'autism' defence based on diagnostic markers rather than individual characteristics. However, it is essential to emphasise that these conclusions come with some crucial caveats. Although this research suggests that autistic individuals may not be

inherently disadvantaged when detecting suspicious behaviour, there does appear to be a small cohort or sub-group of autistic adults with pronounced ToM difficulties who may indeed be particularly vulnerable. It also does not discount the possibility that other characteristics associated with autism, such as restricted interests, sensory sensitivities, or desire for social affiliations could potentially render someone more vulnerable to criminal involvement. In addition, I did not explore the interaction between autism and co-occurring conditions such as intellectual disability or other mental health concerns. Thus, a more nuanced understanding of the relationship between autism and criminal behaviour is essential. Specifically, empirical investigation of potential interactions between the aforementioned factors, rather than a general assumption that a range of characteristics that might heighten vulnerability characterise all autistic adults, is necessary.

References

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/https://doi.org/10.1016/S0885-2014(00)00014-9

Allely, C. S. (2019). Firesetting and arson in individuals with Autism Spectrum Disorder: A systematic PRISMA review. *Journal of Intellectual Disabilities and Offending Behaviour, 10*(4), 89-101. https://doi.org/10.1108/Jidob-11-2018-0014

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.).

Astle, D. E., & Fletcher-Watson, S. (2020). Beyond the core-deficit hypothesis in developmental disorders. *Current Directions in Psychological Science, 29*(5), 431-437. https://doi.org/10.1177/0963721420925518

Baddeley, A., & Crawford, J. (2012). *Spot the word*. Pearson Assessment.

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

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson Rosenberg, C., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L.-C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., Pettygrove, S., Constantino, J. N., Vehorn, A., Shenouda, J., Hall-Lande, J., Van Naarden Braun, K., & Dowling, N. F. (2018). Prevalence of Autism Spectrum Disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014.

Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002), 67(6), 1-23. https://doi.org/10.15585/mmwr.ss6706a1

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development, 2004,* 37–49. https://doi.org/10.1002/cd.96

Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes". *Intelligence, 44*, 78-92. https://doi.org/https://doi.org/10.1016/j.intell.2014.03.001

Baron-Cohen, S. (2001). Theory of mind and autism: A review. *International Review of Research in Mental Retardation, 23*, 169-184. <Go to ISI>://WOS:000165696200007

Baron-Cohen, S. (2008). Autism, hypersystemizing, and truth. *The Quarterly Journal of Experimental Psychology, 61*(1), 64-75. https://doi.org/10.1080/17470210701508749

Baron-Cohen, S. (1988). An assessment of violence in a young man with Asperger's syndrome. *Journal of Child Psychology and Psychiatry, 29*(3), 351-360.

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/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. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

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

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

Bellesi, G., Vyas, K., Jameel, L., & Channon, S. (2018). Moral reasoning about everyday situations in adults with autism spectrum disorder. *Research in Autism Spectrum Disorders, 52*, 1-11. https://doi.org/10.1016/j.rasd.2018.04.009

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

Brewer, N., Lucas, C. A., Lim, A., & Young, R. L. (2023). Detecting dodgy behaviour: The role of autism, autistic traits and theory of mind. *Autism, 27*(4), 1026-1035. https://doi.org/10.1177/13623613221125564

Brewer, N., Wei Ying, A. B., 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/10.1016/j.jarmac.2017.09.006

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*(1), 11-30. https://doi.org/10.1037/1076-898X.12.1.11

Brewer, N., & Young, R. L. (2015). *Crime and autism spectrum disorder: Myths and mechanisms*. Jessica Kingsley Publishers.

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring theory of mind in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(7), 1927-1941. https://doi.org/10.1007/s10803-017-3080-x

Brewer, N., Young, R. L., Norris, J. E., Maras, K., Michael, Z., & Barnett, E. (2022). A quick measure of Theory of Mind in autistic adults: Decision accuracy, latency and self-awareness. *Journal of Autism and Developmental Disorders, 52*(6), 2479-2496. https://doi.org/10.1007/s10803-021-05166-7

Brewer, N., Zoanetti, J., & Young, R. L. (2019). Convergent Validity of the A-ToM (Adult Theory of Mind) Test for Individuals With Autism Spectrum Disorder. *Journal of Psychoeducational Assessment, 37*(6), 797-802. https://doi.org/10.1177/0734282918787433

Brown, G. M. (2006). Degrees of doubt: Legitimate, real and fake qualifications in a global market. *Journal of Higher Education Policy and Management, 28*(1), 71-79. https://doi.org/10.1080/13600800500440789

Browning, A., & Caulfield, L. (2011). The prevalence and treatment of people with Asperger's Syndrome in the criminal justice system. *Criminology & Criminal Justice, 11*(2), 165-180. https://doi.org/10.1177/1748895811398455

Buon, M., Dupoux, E., Jacob, P., Chaste, P., Leboyer, M., & Zalla, T. (2013). The role of causal and intentional judgments in moral reasoning in individuals with high functioning autism. *Journal of Autism and Developmental Disorders, 43*(2), 458-470. https://doi.org/10.1007/s10803-012-1588-7

Cashin, A., & Newman, C. (2009). Autism in the criminal justice detention system: A review of the literature. *Journal of Forensic Nursing, 5*(2), 70-75. https://doi.org/10.1111/j.1939-3938.2009.01037.x

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12*(1), 53-81. https://doi.org/10.1146/annurev-clinpsy-021815-093623

Chandler, R. J., Russell, A., & Maras, K. L. (2019). Compliance in autism: Self-report in action. *Autism, 23*(4), 1005-1017. https://doi.org/10.1177/1362361318795479

Cialdini, R. B., Vincent, J. E., Lewis, S. K., Catalan, J., Wheeler, D., & Darby, B. L. (1975). Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology, 31*(2), 206–215. https://doi.org/10.1037/h0076284

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37*(3/4), 256-266. https://doi.org/10.2307/2332378

Cochran, W. G. (1954). Some methods for strengthening the common χ 2 tests. *Biometrics, 10*(4), 417. https://doi.org/10.2307/3001616

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Collins, J., Horton, K., Gale-St Ives, E., Murphy, G., & Barnoux, M. (2023). A systematic review of autistic people and the criminal justice system: An update of King and Murphy (2014). *Journal of Autism and Developmental Disorders*, *53*(8), 3151–3179. https://doi.org/10.1007/s10803-022-05590-3

Dahling, J. J., Whitaker, B. G., & Levy, P. E. (2009). The development and validation of a new Machiavellianism scale. *Journal of Management, 35*(2), 219-257. https://doi.org/10.1177/0149206308318618

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., Hickie, I., & Guastella, A. J. (2018). Autism spectrum disorders: a meta-analysis of executive function. *Molecular Psychiatry, 23*(5), 1198-1204. https://doi.org/10.1038/mp.2017.75

Demurie, E., De Corel, M., & Roeyers, H. (2011). Empathic accuracy in adolescents with autism spectrum disorders and adolescents with attention-deficit/hyperactivity disorder. *Research in Autism Spectrum Disorders, 5*(1), 126-134. https://doi.org/https://doi.org/10.1016/j.rasd.2010.03.002

Doherty, M. (2008). *Theory of mind: How children understand others' thoughts and feelings*. Psychology Press.

Dubreucq, J., Haesebaert, F., Plasse, J., Dubreucq, M., & Franck, N. (2022). A systematic review and meta-analysis of social skills training for adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 52*(4), 1598-1609. https://doi.org/10.1007/s10803-021-05058-w

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fisher, M. H., Baird, J. V., Currey, A. D., & Hodapp, R. M. (2016). Victimisation and social vulnerability of adults with intellectual disability: A review of research extending beyond Wilson and Brewer. *Australian Psychologist, 51*(2), 114-127. https://doi.org/10.1111/ap.12180

Fisher, M. H., Moskowitz, A. L., & Hodapp, R. M. (2013). Differences in social vulnerability among individuals with autism spectrum disorder, Williams syndrome, and Down

syndrome. *Research in Autism Spectrum Disorders, 7*(8), 931-937. https://doi.org/10.1016/j.rasd.2013.04.009

Freckelton, I. (2013). Autism spectrum disorder: Forensic issues and challenges for mental health professionals and courts. *Journal of Applied Research in Intellectual Disabilities, 26*(5), 420-434. https://doi.org/https://doi.org/10.1111/jar.12036

Freckelton, I., & List, D. (2009). Asperger's Disorder, criminal responsibility and criminal culpability. *Psychiatry, Psychology and Law, 16*(1), 16-40. https://doi.org/10.1080/13218710902887483

Freeman, G., & Halton, J. H. (1951). Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika, 38*(1/2), 141-149.

Frith, U. (1994). Autism and theory of mind in everyday life. *Social Development, 3*(2), 108-124. https://doi.org/10.1111/j.1467-9507.1994.tb00031.x

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/https://doi.org/10.1111/j.1469-7610.2004.00262.x

Frith, U., Hill, E. L., Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 345-360. https://doi.org/doi:10.1098/rstb.2002.1202

Fuller, E. A., & Kaiser, A. P. (2020). The effects of early intervention on social communication outcomes for children with Autism Spectrum Disorder: A meta-analysis. *Journal of Autism and Developmental Disorders, 50*(5), 1683-1700. https://doi.org/10.1007/s10803-019-03927-z

Gal, E., Landes, E., & Katz, N. (2015). Work performance skills in adults with and without high functioning autism spectrum disorders (HFASD). *Research in Autism Spectrum Disorders, 10*, 71-77. https://doi.org/https://doi.org/10.1016/j.rasd.2014.10.011

George, M. S., Teunisse, A. K., & Case, T. I. (2020). Gotcha! Behavioural validation of the Gullibility Scale. *Personality and Individual Differences, 162*, 110034. https://doi.org/https://doi.org/10.1016/j.paid.2020.110034

Gernsbacher, M. A., & Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a theory of mind. *Archives of Scientific Psychology, 7*(1), 102-118. https://doi.org/10.1037/arc0000067

Grant, T., Furlano, R., Hall, L., & Kelley, E. (2018). Criminal responsibility in autism spectrum disorder: A critical review examining empathy and moral reasoning. *Canadian Psychology/Psychologie Canadienne, 59*(1), 65-75. https://doi.org/10.1037/cap0000124

Gray, K., Jenkins, A. C., Heberlein, A. S., & Wegner, D. M. (2011). Distortions of mind perception in psychopathology. *Proceedings of the National Academy of Sciences, 108*(2), 477-479. https://doi.org/10.1073/pnas.1015493108

Grieve, R., & Mahar, D. (2013). Can social intelligence be measured? Psychometric properties of the Tromsø Social Intelligence Scale—English Version. *The Irish Journal of Psychology, 34*, 1-12. https://doi.org/10.1080/03033910.2012.737758

Griffiths, S., Allison, C., Kenny, R., Holt, R., Smith, P., & Baron-Cohen, S. (2019). The Vulnerability Experiences Quotient (VEQ): A study of vulnerability, mental health and life satisfaction in autistic adults. *Autism Research, 12*(10), 1516-1528. https://doi.org/https://doi.org/10.1002/aur.2162

Happé, F. G. (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.

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843-855. https://doi.org/10.2307/1131954

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders, 30*(3), 225-236. https://doi.org/10.1023/A:1005544518785

Hendricks, D. (2010). Employment and adults with autism spectrum disorders: Challenges and strategies for success. *Journal of Vocational Rehabilitation, 32*, 125-134. https://doi.org/10.3233/JVR-2010-0502

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): a study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38*(8), 1555-1566. https://doi.org/10.1007/s10803-008-0538-x

Howlin, P. (2004). *Autism and Asperger syndrome: Preparing for adulthood*. Routledge.

Jolliffe, T., & Baron-Cohen, S. (1999). The Strange Stories Test: A replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders, 29*(5), 395-406. https://doi.org/10.1023/A:1023082928366

Jones, S. C., & Harwood, V. (2009). Representations of autism in Australian print media. *Disability & Society, 24*(1), 5-18. https://doi.org/10.1080/09687590802535345

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1304-1316.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Karoğlu, N., Ferguson, H. J., & Ciardha, C. O. (2021). Theory of mind in offending: A systematic review. *Trauma, Violence, and Abuse*, *6*, Article 1013143. https://doi.org/10.1177/15248380211013143

Ke, F., Moon, J., & Sokolikj, Z. (2022). Virtual reality–based social skills training for children with autism spectrum disorder. *Journal of Special Education Technology, 37*(1), 49-62. https://doi.org/10.1177/0162643420945603

Kellaher, D. C. (2015). Sexual behavior and autism spectrum disorders: An update and discussion. *Current Psychiatry Reports, 17*(4), Article 25. https://doi.org/10.1007/s11920-015-0562-4

Kimhi, Y. (2014). Theory of mind abilities and deficits in autism spectrum disorders. *Topics in Language Disorders, 34*(4), 329-343. https://doi.org/10.1097/tld.0000000000000033

King, C., & Murphy, G. H. (2014). A systematic review of people with autism spectrum disorder and the criminal justice system. *Journal of Autism and Developmental Disorders, 44*(11), 2717-2733. https://doi.org/10.1007/s10803-014-2046-5

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders, 31*(1), 29-36. https://doi.org/10.1023/A:1005657512379

Kourtesis, P., Kouklari, E. C., Roussos, P., Mantas, V., Papanikolaou, K., Skaloumbakas, C., & Pehlivanidis, A. (2023). Virtual reality training of social skills in adults with autism spectrum disorder: An examination of acceptability, usability, user experience, social skills, and executive functions. *Behavioural Sciences (Basel, Swizerland), 13*(4), Article 336. https://doi.org/10.3390/bs13040336

Kumagami, T. (2006). Characteristics of juvenile court cases with pervasive developmental disorder. *Seishin shinkeigaku zasshi = Psychiatria et neurologia Japonica, 108*(4), 327-336. http://europepmc.org/abstract/MED/16761693

Leslie, A. M., Mallon, R., & DiCorcia, J. A. (2006). Transgressors, victims, and cry babies: Is basic moral judgment spared in autism? *Social Neuroscience, 1,* 270–283. https://doi.org/10.1080/17470910600992197

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433-442. https://doi.org/10.3758/s13428-016-0727-z

Lodi-Smith, J., Rodgers, J. D., Cunningham, S. A., Lopata, C., & Thomeer, M. L. (2019). Meta-analysis of Big Five personality traits in autism spectrum disorder. *Autism, 23*(3), 556-565. https://doi.org/10.1177/1362361318766571

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

Mackinnon, A., & Christensen, H. (2007). An investigation of the measurement properties of the Spot-the-Word test in a community sample. *Psychological Assessment, 19*(4), 459.

Maïano, C., Normand, C. L., Salvas, M.-C., Moullec, G., & Aimé, A. (2016). Prevalence of school bullying among youth with autism spectrum disorders: A systematic review and meta-analysis. *Autism Research, 9*(6), 601-615. https://doi.org/https://doi.org/10.1002/aur.1568

Mandelberg, J., Laugeson, E. A., Cunningham, T. D., Ellingsen, R., Bates, S., & Frankel, F. (2014). Long-term treatment outcomes for parent-assisted social skills training for adolescents with autism spectrum disorders: The UCLA PEERS Program. *Journal of Mental Health Research in Intellectual Disabilities, 7*(1), 45-73. https://doi.org/10.1080/19315864.2012.730600

Maras, K. L., & Bowler, D. M. (2012). Brief report: Suggestibility, compliance and psychological traits in high-functioning adults with autism spectrum disorder. *Research in Autism Spectrum Disorders, 6*(3), 1168-1175. https://doi.org/10.1016/j.rasd.2012.03.013

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities, 30*(6), 1107-1114. https://doi.org/https://doi.org/10.1016/j.ridd.2009.06.003

Mawson, D., Grounds, A., & Tantam, D. (1985). Violence and Asperger's syndrome: A case study. *The British Journal of Psychiatry, 147*(5), 566-569. https://doi.org/10.1192/bjp.147.5.566

Mazurek, M. O. (2014). Loneliness, friendship, and well-being in adults with autism spectrum disorders. *Autism, 18*(3), 223-232. https://doi.org/10.1177/1362361312474121

McHugh, M. L. (2012). Interrater reliability: the Kappa statistic. *Biochemia Medica, 22*(3), 276-282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

Mouridsen, S. E. (2012). Current status of research on autism spectrum disorders and offending. *Research in Autism Spectrum Disorders, 6*(1), 79-86. https://doi.org/https://doi.org/10.1016/j.rasd.2011.09.003

Nilsson, K. K., & de López, K. J. (2016). Theory of Mind in children with specific language impairment: A systematic review and meta-analysis. *Child Development, 87*(1), 143-153. https://doi.org/https://doi.org/10.1111/cdev.12462

North, A. S., Russell, A. J., & Gudjonsson, G. H. (2008). High functioning autism spectrum disorders: an investigation of psychological vulnerabilities during interrogative interview. *The Journal of Forensic Psychiatry & Psychology, 19*(3), 323-334. https://doi.org/10.1080/14789940701871621

O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of Autism and Developmental Disorders, 39*(6), 916-928. https://doi.org/10.1007/s10803-009-0699-2

Oliveira, C., Simões de Almeida, R., & Marques, A. (2021). Virtual reality in social skills training programs for people with schizophrenia: A systematic review and focus group. *British Journal of Occupational Therapy, 84*(9), 571-581. https://doi.org/10.1177/03080226211011391

Orsmond, G. I., Krauss, M. W., & Seltzer, M. M. (2004). Peer relationships and social and recreational activities among adolescents and adults with autism. *Journal of Autism and Developmental Disorders, 34*(3), 245-256.

O'Sullivan, O. P. (2018). Autism spectrum disorder and criminal responsibility: historical perspectives, clinical challenges and broader considerations within the criminal justice

system. *Irish Journal of Psychological Medicine, 35*(4), 333-339. https://doi.org/10.1017/ipm.2017.13

Ozonoff, S. (1995). Executive functions in autism. In E. Schopler & G. B. Mesibov (Eds.), *Learning and Cognition in Autism* (pp. 199-219). Springer US. https://doi.org/10.1007/978-1-4899-1286-2_11

Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of Intellectual Disability Research, 46*(5), 430-443. https://doi.org/https://doi.org/10.1046/j.1365-2788.2002.00425.x

Peñuelas-Calvo, I., Sareen, A., Sevilla-Llewellyn-Jones, J., & Fernández-Berrocal, P. (2019). The "Reading the Mind in the Eyes" test in autism-spectrum disorders comparison with healthy controls: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders, 49*(3), 1048-1061. https://doi.org/10.1007/s10803-018-3814-4

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*(4), 1643-1662. https://doi.org/10.3758/s13428-021-01694-3

Pinsker, D. M., McFarland, K., & Stone, V. E. (2011). The Social Vulnerability Scale for older adults: An exploratory and confirmatory factor analytic study. *Journal of Elder Abuse & Neglect, 23*(3), 246-272. https://doi.org/10.1080/08946566.2011.584049

Pinsker, D. M., Stone, V., Pachana, N., & Greenspan, S. (2006). Social Vulnerability Scale for older adults: Validation study. *Clinical Psychologist, 10*(3), 109-119. https://doi.org/https://doi.org/10.1080/13284200600939918

Ponnet, K., Buysse, A., Roeyers, H., & De Clercq, A. (2008). Mind-reading in young adults with ASD: Does structure matter? *Journal of Autism and Developmental Disorders, 38*(5), 905-918. https://doi.org/10.1007/s10803-007-0462-5

Quesque, F., & Rossetti, Y. (2020). What do Theory-of-Mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science, 15*(2), 384-396. https://doi.org/10.1177/1745691619896607

R v Middleton (2023) ACTSC 50

Railey, K. S., Love, A. M. A., & Campbell, J. M. (2020). A scoping review of autism spectrum disorder and the criminal justice system. *Review Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s40489-020-00203-0

Roeyers, H., Buysse, A., Ponnet, K., & Pichal, B. (2001). Advancing advanced mind-reading tests: Empathic accuracy in adults with a pervasive developmental disorder. *Journal of Child Psychology and Psychiatry, 42*(2), 271-278. https://doi.org/https://doi.org/10.1111/1469-7610.00718

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*, 651-665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35*, 1-7. https://doi.org/10.1037/0003-066X.35.1.1

Rutten, A. X., Vermeiren, R. R. J. M., & Van Nieuwenhuizen, C. (2017). Autism in adult and juvenile delinquents: a literature review. *Child and Adolescent Psychiatry and Mental Health, 11*(45), 1-12. https://doi.org/10.1186/s13034-017-0181-4

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition, 133*(1), 43-47. https://doi.org/https://doi.org/10.1016/j.cognition.2014.05.016

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763-1768. https://doi.org/10.1213/ane.0000000000002864

Sellers, W. (2018, 8 November). Timeline of a murder: How autistic man, 19, drank heavily before stalking Eurydice Dixon for more than an hour - as its revealed distressing details were kept from her familly. *Daily Mail Australia*. https://www.dailymail.co.uk/news/article-6366995/How-autistic-man-19-drank-heavily-stalking-Eurydice-Dixon-hour.html

Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist, 18*(2), 108-113. https://doi.org/10.1177/1073858410397208

Shulman, C., Guberman, A., Shiling, N., & Bauminger, N. (2012). Moral and social reasoning in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*(7), 1364-1376. https://doi.org/10.1007/s10803-011-1369-8

Silvera, D., Martinussen, M., & Dahl, T. I. (2001). The Tromsø Social Intelligence Scale, a self-report measure of social intelligence. *Scandinavian Journal of Psychology, 42*(4), 313-319.

Soares, E. E., Bausback, K., Beard, C. L., Higinbotham, M., Bunge, E. L., & Gengoux, G. W. (2021). Social skills training for autism spectrum disorder: A meta-analysis of in-person and technological interventions. *Journal of Technology in Behavioral Science, 6*(1), 166-180. https://doi.org/10.1007/s41347-020-00177-0

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2010). Theory of Mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders, 40*(3), 280-289. https://doi.org/10.1007/s10803-009-0860-y

Spenser, K. A., Betts, L. R., & Das Gupta, M. (2015). Deficits in Theory of Mind, empathic understanding and moral reasoning: a comparison between young offenders and non-offenders. *Psychology, Crime & Law, 21*(7), 632-647. https://doi.org/10.1080/1068316x.2015.1028542

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*(5), 640-656. https://doi.org/10.1162/089892998562942

Strunz, S., Schermuck, C., Ballerstein, S., Ahlers, C. J., Dziobek, I., & Roepke, S. (2017). Romantic relationships and relationship satisfaction among adults with Asperger syndrome and high-functioning autism. *Journal of Clinical Psychology, 73*(1), 113-125. https://doi.org/https://doi.org/10.1002/jclp.22319

Takeda, T., Kasai, K., & Kato, N. (2007). Moral judgment in high-functioning pervasive developmental disorders. *Psychiatry and Clinical Neurosciences, 61*(4), 407-414. https://doi.org/https://doi.org/10.1111/j.1440-1819.2007.01678.x

Taylor, J. L., Henninger, N. A., & Mailick, M. R. (2015). Longitudinal patterns of employment and postsecondary education for adults with autism and average-range IQ. *Autism, 19*(7), 785-793. https://doi.org/10.1177/1362361315585643

Teunisse, A. K., Case, T. I., Fitness, J., & Sweller, N. (2020). I should have known better: Development of a self-report measure of gullibility. *Personality and Social Psychology Bulletin, 46*(3), 408-423.

Tobacyk, J. (2004). A revised Paranormal Belief Scale. *International Journal of Transpersonal Studies, 23*(1), 94-98. https://doi.org/10.1037/t14015-000

Wechsler, D. (2008). *Wechsler Aadult Intelligence Scale–Fourth Edition (WAIS–IV)*. Pearson.

Wiedmaier, B. (2017). Phi Coefficient. In M. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods* (Vol. 4). SAGE Publications, Inc. https://doi.org/https://doi.org/10.4135/9781483381411

Weiss, J. A., & Fardella, M. A. (2018). Victimization and perpetration experiences of adults with autism. *Front Psychiatry, 9*, 1-10. https://doi.org/10.3389/fpsyt.2018.00203

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.

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. *Autism Research, 11*(8), 1129-1137. https://doi.org/10.1002/aur.1962

Wilson, C., & Brewer, N. (1992). The incidence of criminal victimisation of individuals with an intellectual disability. *Australian Psychologist, 27*(2), 114-117. https://doi.org/10.1080/00050069208257591

Wolf, G. (2021). Growing enlightenment: Sentencing offenders with autism spectrum disorder in Australia. *University of New South Wales Law Journal, 44*(4), 1701-1738. https://doi.org/10.53637/RWWG5986

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin 124*(3), 283-307. https://doi.org/10.1037/0033-2909.124.3.283

Young, N., Anderson, B., & Radutzky, M. (2012, 17 December). *60 Minutes Reports: Tragedy in Newtown*. CBS News. Retrieved 15 July from https://www.cbsnews.com/news/60-minutes-reports-tragedy-in-newtown/

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

Yu, Y., Bradley, C. C., Boan, A. D., Charles, J. M., & Carpenter, L. A. (2021). Young adults with autism spectrum disorder and the criminal justice system. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-020-04805-9

Yuan, S. N. V., & Ip, H. H. S. (2018). Using virtual reality to train emotional and social skills in children with autism spectrum disorder. *London Journal of Primary Care, 10*(4), 110-112. https://doi.org/10.1080/17571472.2018.1483000

Yuspeh, R. L., & Vanderploeg, R. D. (2000). Spot-the-Word: A measure for estimating Ppremorbid intellectual functioning. *Archives of Clinical Neuropsychology, 15*(4), 319-326. https://doi.org/https://doi.org/10.1016/S0887-6177(99)00020-7

Zalla, T., Sav, A.-M., Stopin, A., Ahade, S., & Leboyer, M. (2009). Faux pas detection and intentional action in Asperger syndrome. A replication on a French sample. *Journal of Autism and Developmental Disorders, 39*(2), 373-382.

Appendix A

Crime Scenario 2 (Andy)

Segment 1:

Andy is dropping his little brother off at football training. He is standing by his car talking to another boy's father, Carl, whom he has known for a number of years.

Carl: Andy, how have you been?

Andy: Oh you know, It's been tough getting work. I've had a few applications knocked back and there's not much else going around at the moment.

Carl: Yeah, I get you. I have some work you could help me out with. It involves driving some people around. You'd get paid by the hour and petrol costs will all be covered. What do you reckon?

Segment 2:

Andy: Yeah mate, that sounds really good.

Carl: Alright well I'll give you a call soon when I've got a job for you.

The next day, Andy is at home when he receives a call. It's from Carl.

Carl: Hey mate, got a pickup for you. I'll give you a cash payment when you're out at footy next.

Segment 3:

Andy: Alright, message me the address.

Andy pulls up at the address Carl gave him. The suburb seems seedy and the house he pulls up in front of is run down with peeling paint. Andy pulls out his phone to call Carl and double check the address when a woman leaves the house and hastily walks toward the car. She's wearing a backpack, untidy hair and ripped jeans.

Segment 4:

The woman jumps into the back of Andy's car. He leans into the back to sweep away some bottles and a jumper.

Andy: Ahh sorry about the mess… Where do you want to go?

Girl: Westfield shopping centre, it's about 10 minutes away.

Andy drives and the woman doesn't say a thing. Andy doesn't know what to say either, so they ride in silence. They arrive at the shopping centre carpark. The woman leans forward and says quietly

Girl: Can you drop me off by the bank?

Segment 5:

Andy drives to the bank. She thanks him and gets out. When he waits and watches for a moment she turns around and glances at him, so he drives off.

Carl calls Andy again in a couple of days. The new address is again in a run-down suburb and a person comes straight out to the car. He has a gym bag and asks to be taken to a liquor store. He doesn't want to talk so they ride in silence.

A week later, the person Andy has to pick up asks to be dropped at the courthouse.

At the next football training, Carl pays Andy in cash as promised.

> **Carl:** Here you go mate. I'll have another pickup for you in the next couple of days so keep your phone handy.

Segment 6:

Carl texts Andy a couple of nights later. It's 11pm and he needs a boy to be picked up from a shopping centre. When Andy arrives, everything is closed but a teenager is waiting. As the boy gets in the car, Andy can tell he has a blood nose.

> **Andy:** Are you ok?
>
> **Boy:** Yeah, don't worry about it.

The requested drop off spot is a dark, glass building close to the city centre. As Andy drops the boy off, he asks:

> **Boy:** Would you mind waiting around for a minute?

Segment 7:

Andy watches him walk toward the building and duck around a corner into the darkness. All of a sudden, an alarm sounds, and the boy comes sprinting back towards the car. A security guard exits the building, shouting, with a phone in one hand. As the boy jumps back in Andy's car, sirens can be heard approaching.

Crime Scenario 3 (Amy)

Segment 1

Amy is chatting with her friends Em and Jake about their plans to move out together in the new year, once they finish year 12. She tells them she will need to get a job before she can afford any of the living expenses. A few days later, while shopping with her dad, Amy sees a request for staff at the local Coles. She walks in and submits an application to the manager. She soon hears she has been hired and begins her training as a check-out-chick.

> **Amy's Boss:** You'll need to hold polite, repetitive conversation with every customer that comes through. Some customers will be pretty chatty and some won't, so you might hear all about a customer's hip surgery one minute and then serve a customer that barely grunts hello.
>
> **Amy:** No problem, I can manage that.

A few months later, Amy is working an afternoon shift when an extremely chatty man comes through the checkout.

> **Man:** You look familiar, young lady. Did you go to a school around here?

Segment 2

Conversation continues while Amy puts the man's items through the checkout. He is buying tinned soups and vegetables, discounted meat and other discounted products.

> **Amy:** I went to St Mark's. Just finished year 12.
>
> **Man:** Ahhh yes, my grandson went to your school in a lower grade. What subjects did you choose? And what are you hoping to do at University?
>
> **Amy:** I did Chemistry, Biology, Physics, English and Maths. I'd really like to do physio, but I don't know which Uni or whether I'll get in. Your total is $28.60 please.

They exchange money and Amy gives change of $1.40. The man looks confused.

> **Man:** I thought I just gave you $40. My change should be more than that.

Segment 3

Amy looks confused.

> **Amy:** I'm so sorry, I thought you'd given me a twenty and a ten.
>
> **Man:** No, it was definitely $40. I only ever take out $60 on grocery day and I have just $20 left.

The man opens his wallet and Amy sees there is nothing but a $20 note remaining.

Segment 4

> **Amy:** I'm sorry, I mustn't have been paying attention.

Amy issues a refund and gives the man $11.40.

Next week, Amy arrives to work the same shift. As she is about to begin serving someone, Amy sees her friend Lyla clock-on at the checkout behind her.

Amy: Let's catch up during lunch break!

Lyla: Sounds good!

Lyla starts working at her station. Amy has just finished giving change and receipt to the woman she is serving when she looks up and sees the man from last week next in line. His eyes widen and he hesitates, then walks to the next checkout over. Amy's checkout is now empty, and he has lined up behind an elderly lady.

Segment 5

Amy shrugs it off. She begins serving her next customer, but he isn't up for a chat, so she works in silence. Amy can overhear conversation between the man and Lyla behind her.

Man: That's great. So, are you planning to do any travelling any time soon?

Lyla: Yep, I have tickets booked for South America at the end of the year. Here's your change.

Man: No... that can't be my change. I gave you two twenties, not a twenty and a ten.

Amy looks over her shoulder to see the man showing the separate compartments in his wallet, just as he had the week before.

Lyla: Oh, I'm so sorry, here you go.

Lyla gives the man an extra $10.

Segment 6

Amy approaches Lyla when the man has left.

Amy: Lyla, that exact same thing happened when I served that man last week. I think we need to report him to management.

Lyla: Really? I knew I had given him the correct change! Hmm. You're right. Let's go find a supervisor.

Crime Scenario 4 (Jia)

Segment 1

Jia is in a university club meeting for international students. They are discussing their plans for after they finish their final year.

> **Friend:** Jia, what are you thinking about doing next year? Your student visa expires soon, right?
>
> **Jia:** Yeah, I have no idea at the moment. My job applications haven't gone too well so I haven't been able to get hold of a work visa. I love being in Australia though, I really don't want to have to leave.

A month later, Jia and her friends are out celebrating because she has just handed in her Honours thesis. During the night, she is anxiously chatting with her friend Samantha about what might come next.

> **Jia:** I don't know what to do about this visa situation. Time's running out and I'm getting really stressed.
>
> **Samantha:** Well, my cousin recently enrolled in an online, part-time graduate diploma program at a Uni in Sydney. It's meant he can stay in Australia on a study visa while he searches for jobs.
>
> **Jia:** Oh, that sounds so good, what's it called?
>
> **Samantha:** Not too sure, I'll give you my cousin's email so he can give you all the details

Segment 2

After receiving an email reply from Samantha's cousin, Jia starts searching the web. The search shows a website that claims the university offers "exceptional education for students around the world wanting to study in Australia". It lists an address in Sydney, but courses were all online anyway. The website lists seven undergraduate programs and nine graduate programs. Jia spots a practical-based business diploma that seems to fit her needs. The course page says that the first semester only involves work experience at a business. If you have prior business experience, then it's possible that no class time is necessary.

Segment 3

Jia emails the course adviser listed on the page, and he responds the next day to reassure her and explain the course in more detail. He also tells her the first semester costs $3450 and needs to be paid upfront in full. Jia speaks to her parents and shows them the course website. They agree to pay half.

Segment 4

Jia enrols in the diploma, pays the fees, and submits the forms to extend her student visa.

She is also working around 20 hours per week at a cafe, and soon gets promoted to manager. She keeps up her course work which is to keep a weekly journal of her work and responsibilities. She

receives an email once a month from her course instructor reminding her to submit journal entries. Jia emails her coordinator asking about enrolment for the following semester. He never replies.

Segment 5

Towards the end of the semester, Jia finds out she has been selected for an amazing volunteer position at the Red Cross where she is trained in disaster counselling. Jia completed a double degree in business and psychology, so the position utilises the skills she learned. That night, she tells her parents she's finally found something she'd like to do for a career.

Jia meets Sally at her Red Cross training and they soon become friends. She is in the same position Jia was last year.

> **Sally:** I'm so worried about what I'll do next year when my visa is up.
>
> **Jia:** Oh, I just submitted my final business report for an online course I've been doing! It's meant that I can stay on a student visa while I look for work. I'll send you the website link.
>
> **Sally:** Oh thanks! That'd be great.

When they see each other again the next week, Sally seems worried.

> **Sally:** Jia, about that course you told me to look at… I couldn't find any information about actual online lectures or classes on the website… Have you been doing any? I couldn't find any staff profiles either... Or a Facebook or Twitter account.

Segment 6

Jia receives an email from her course coordinator with instructions about how to enrol for next semester. She leaves it for now, because what Sally had told her has her concerned. After all, she hasn't attended a single class.

A week later, the police show up at Jia's doorstep and ask her to come in for questioning.

Non-Crime Scenario 1 (Casey)

Segment 1:

Casey has been working the same office job for five years. One day, she catches up with her old friend, Teresa, and they discuss how their lives have changed since university.

> **Casey:** I can't stand this job. I hate being in the city every day, working under artificial lighting, staring at a screen for hours, barely exercising... I feel so unhealthy.
>
> **Teresa:** You need to get outside more. Come and join my cycling group on weekends. It's so fun and it'll be really good for you.
>
> **Casey:** You know what, why not? Sounds interesting.

Casey starts going along to the cycling events on weekends, and soon begins making friends. One of the guys in the group, Ian, is a cycling coach and lends Casey a bike to use. She learns several new routes around the city and begins to really enjoy riding.

Soon, Casey is cycling every weekend and starts to look for a bike of her own. Ian gives her some tips on what to buy.

> **Ian:** You want to look for a road-specific bike but try to avoid the ones with fixed gears. Good road bikes can be pretty expensive, so look out for dodgy deals. A second-hand one should do the trick while you're starting out though.

Casey spends her lunch breaks at work reading cycling forums and scrolling through buy-and-sell websites. Finally, one Friday, she finds a bike that looks perfect — it's a recent model, used but still in very good condition.

She messages the seller, and they agree to meet on Sunday. On Saturday she goes cycling as usual. She brings the borrowed bike and is dismounting at the meeting spot when Ian arrives, in a furious mood.

Segment 2:

> **Casey:** You OK Ian?
>
> **Ian:** Someone broke into my shed at home last Thursday night and stole my bike. I've reported it to the police but there were no witnesses and no real leads, so they aren't hopeful of recovering it. I just can't believe it.

Segment 3:

> **Casey:** Oh man, I'm so sorry to hear that. I know you really love that bike.

On Sunday, Casey drives to pick up the bike. The seller is two suburbs over from where she lives. He is a tall, lean middle-aged man who takes her through the house to the back veranda where the bike is propped up against a chair.

> **Seller:** She's in great condition. I bought it for my girlfriend recently but after a few rides she decided she didn't like it anymore. Now I need the money, so I'm selling.

Segment 4:

Casey looks again at the bike – it looks great and is reasonably priced, so she buys it.

The next Saturday she hauls her new bike up onto her roof racks, along with the bike borrowed from Ian, and drives to the meeting spot. She's excited to show it off to her fellow club members, and as she walks it over to the group, a few of them come up to admire it and congratulate her on her first bike. One of them says that it looks weirdly familiar but can't place why.

Segment 5:

Ian turns up a few minutes late with one of his old bikes. When he sees Casey's new bike his jaw drops. He starts shouting and begins to run towards her. Others in the club tell him to quiet down since it's still early morning, so Casey can't hear what he is trying to say until he gets closer.

Ian: That's my bike! Look at the paint near the front wheel. There'll be a big scratch from where I fell off crossing some tram tracks.

Segment 6:

Casey is still shocked about the whole situation but shifts the bike so that they can both see the front wheel. There is no scratch.

> **Ian:** Oh… Oh no. I'm so sorry. It must just be the same model. I assumed it must have been mine, I'm sorry, I was too quick to judge.

Casey lets out a big sigh of relief. She feels sorry for Ian but is happy her bike isn't stolen property.

Non-Crime Scenario 3 (Derek)

Segment 1:

Derek is studying human movement at university. After submitting an assignment, Derek and his friends are having some drinks at one of their share-houses. During the night, one of the newer girls to the group, Nicole, approaches Derek for a chat.

> **Nicole:** hey Derek, can I ask you a favour?
>
> **Derek:** Yeah sure, what's up?
>
> **Nicole:** I'm heading interstate for a while and don't really feel comfortable leaving some of my valuables at home while I'm gone. My area has a pretty high rate of home break-ins and my room-mates are kind of dodgy so I don't really trust them. You live pretty close, so how would you feel about holding onto some stuff for me while I'm gone? It'll only be for a couple of weeks.

Segment 2:

> **Derek:** Yeah, I can do that for you.
>
> **Nicole:** Thanks, I know that I can trust you. I'll bring the stuff over tomorrow afternoon.
>
> **Derek:** Sure.

The next day, Nicole arrives at Derek's house with her belongings.

> **Nicole:** There's a few boxes for you to hold onto, is that ok?

Segment 3:

> **Derek:** All good, I have enough room in my garage.
>
> **Nicole:** Thanks so much for doing this for me. I'll message you when my holiday is over.

Derek takes the boxes to his garage for storage. They are all taped up so he can't see what's inside.

Segment 4:

A couple of weeks pass, and Derek notices it's now the date Nicole was due to be back from her holiday. He waits for her message but a few days later he still hasn't heard from her.

Segment 5:

Derek messages Nicole and waits for a response. A few minutes later, she texts back.

> **Nicole:** I'm so sorry Derek, I *completely* forgot about leaving my things at your place! I'll come over right away to pick them up.

Within 20 minutes Nicole is at Derek's house, packing her belongs in her car and giving him some travel souvenirs for his trouble.

Non-crime scenario 1 (Charlie)

Segment 1:

Charlie is working as a receptionist at a grand old hotel. His co-worker, Eleanor, makes a mistake and asks for his help.

> **Eleanor:** Ahh Charlie, I've accidentally made a double booking for a room for this weekend. Can you fix it for me? I can't figure out how…

Charlie sighs.

> **Charlie:** Sure. Where's the booking? I'll have to log into the system under my name to fix it up. You need to be more careful though, the guests receive notifications as soon as you make a booking so it's a bit of a process to fix it.

A guest approaches the counter. Over his shoulder, Charlie spots a young woman who doesn't look like the regular clientele.

Segment 2:

> **Rich guest:** Excuse me I'd like to check in.

While he would prefer to check the guest in himself, Charlie thinks he should leave this job to Eleanor so that he can quickly fix her pevious mistake.

> **Charlie:** No problem sir, Eleanor here will help you out.

Eleanor begins to help the man and Charlie looks again at the young woman. She is dressed in jeans and a black t-shirt and her hair is pulled into a messy ponytail. She doesn't approach the front desk but walks around the foyer looking at various paintings. Charlie doesn't take much more notice as she might just be visiting a guest.

The young woman admires a painting on the wall and moves around to look at the lift and nearby brochures. Charlie glances up briefly before returning to fixing Eleanor's mistake.

Charlie finally finishes on the computer. He sighs and looks up. The woman is still there, peeking into the hotel's bar and restaurant. She pulls out a small notebook from her jeans pocket and starts taking notes.

Segment 3:

Charlie frowns, straightens his tie and walks over to the woman. He politely asks:

> **Charlie:** Sorry Miss, can I help you with something?

The woman is startled and quickly tucks away her notepad.

> **Charlie:** You don't look like our regular kind of guest.

Young woman: Ah… I'm sorry… What are your guests usually like?

Charlie: Well, for the most part, they're incredibly wealthy. Often highly regarded or influential in their field of work. They come from all over the globe.

Young woman: How long has the hotel been open? What's it like to work here? And what are the regular dinner hours of the bar?

Segment 4:

Charlie: The hotel opened in 1880 and has since been one of the premier accommodation destinations in Australia. I love working here. The clientele can be demanding but that's what comes with status, I guess… Oh and dinner is generally served between 5 and 10pm.

Young woman: Ok, interesting… What are staff shift changes like? And your managers, how often are they around?

Segment 5:

Charlie looks concerned. The young woman notices.

Young woman: Oh, I'm not a criminal I promise… next week I have an interview to become a chef in the kitchens but I'm really nervous about it, so I wanted to find out more.

Segment 6:

Charlie goes back to work. Next week he sees the girl walk past with the head chef to begin her interview. He later hears she's been hired.

Appendix B

Coding Protocol

Behaviour Codes *(What would you do if you were X in this situation?)*
More than one code may be applicable

(1) Comply with request/accept offer/go ahead
- The participant agrees when they are asked to do something by another character
  - "Simple request, I would agree to wait"
  - "I would immediately look at the bike to see if the paint scratch was there as Ian insisted"
- The participant indicates that they will take up an offer from another character
  - "I would agree to store the files"
  - "Would jump at the opportunity to earn some money"
- The participant decides to 'go ahead' and continue down the path of the story. Generally relevant in the segments after an offer or request has been made, or when no other character is involved
  - "I would sign up for the program"
  - "I would put the boxes on a shelf in the garage and forget about them"

(2) Take action
- The participant indicates that they would do something other than continue along the path of the story, such as involving another character, but which does not necessarily remove them from the situation
  - "I would open a box just to see what it is"
  - "Go to her house to see if she has returned"
  - "Start looking for another job"
  - "Confirm what the images are"
  - "Recount the money"

(3) Deny request/decide against/decline offer
- The participant indicates that they will not do what is asked of them by another character
  - "I would tell him I cannot do that"
  - "I would politely decline"
- The participant indicates that they would turn down an offer made to them
  - "say thanks, but no thanks"
  - "I would say no to the offer"
- The participant decides not to continue with (when no offer/request is made)
  - "If I was Jia I would probably not do this"
  - "I would still be uncomfortable and tell him he needs to find other storage"

(4) Extricate
- The participant indicates they would do something which actively removes them from the situation. This does not include contacting authorities.

- "I would call the manager to confirm that I may have made a mistake"
- "Tell her you got lost and are at the wrong house"
- "I would stop loading the files and delete the others" – (just stopping the download is not enough in this case, code that as a 3)
- "I would explain to Ian how I bought the bike"

(5) Conversational/socially acceptable

- The participant responds in a way that would be considered a conversational or socially acceptable response. This requires them to say something.
  - "I would laugh a little and introduce myself"
  - "I would explain the class structure to Sally"
  - "Ask Ian what's wrong"
  - "I would share some of the information I'd discovered searching for used bikes"

(6) Seek clarification/further information

- The participant indicates that they would seek further information, or would like something clarified, before making a decision. This is generally relevant any time they indicate they would ask a question.
  - "I would ask what sort of valuables she was talking about"
  - "Ask why cash is the only option"
  - "I would ask my friend why he has so many photos of kids"
  - "I would do more research into the program"

(7) Continue as were/no action taken

- The participant indicates that they would not take any real action in that situation, or would continue with the same activity as before
  - "I would be starting to freak out, since I am almost done with the course work"
  - "I would greet the guests and get them checked in"
  - "I would be curious about her, but wouldn't do anything with the boxes"
  - "I would stay in the car"

(8) Involve authorities

- The participant indicates that they would call authorities (such as police). This does <u>not</u> include making contact with a manager or supervisor.
  - "I would immediately contact the authorities and hand over the hard drive to them, while also giving them the information on who I got the hard drive from and where they can be located"
  - "At this point I would contact the Australian authorities about this issue and hand them the information to ensure that I am not being scammed"

(9) Other

- Any response that does not readily fit into these categories

Suspicion Codes *(Why would you do that?)*

(0) No reference to suspicion
- The participant provides an explanation that has no reference to suspicion or is irrelevant
    - "It seems like something that may be a good fit"
    - "I wouldn't want to upset him and cause a scene"
    - "I wouldn't want to drive much further and use more gas when she originally wanted to go to the mall"

(1) Reference to suspicion (indecisive)
- The participant refers to suspicion but does not explicitly suggest that they think the situation is suspicious (or not suspicious). Language such as "weird" and "strange" goes here.
    - "don't be rude but don't volunteer info - you can't be too trusting"
    - "I want to make sure I can get my money back if I don't qualify or something happens"
    - "I would want to know if it actually was his bike"
    - "Could be nothing or could be a thief"

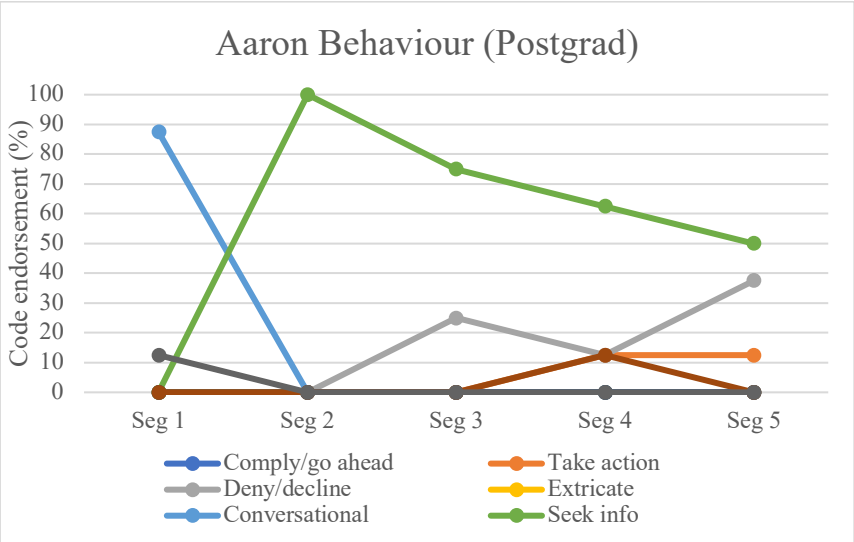
(2) Suspicious
- The participant clearly states the situation is suspicious (but does not necessarily need to use the word suspicious)
    - "I would be starting to think I'm getting ripped off"
    - "Based on the man's behaviour, I assume that he's trying to scam money out of cashiers"
    - "sounds shady, not legitimate"
    - "it just sounds fishy"

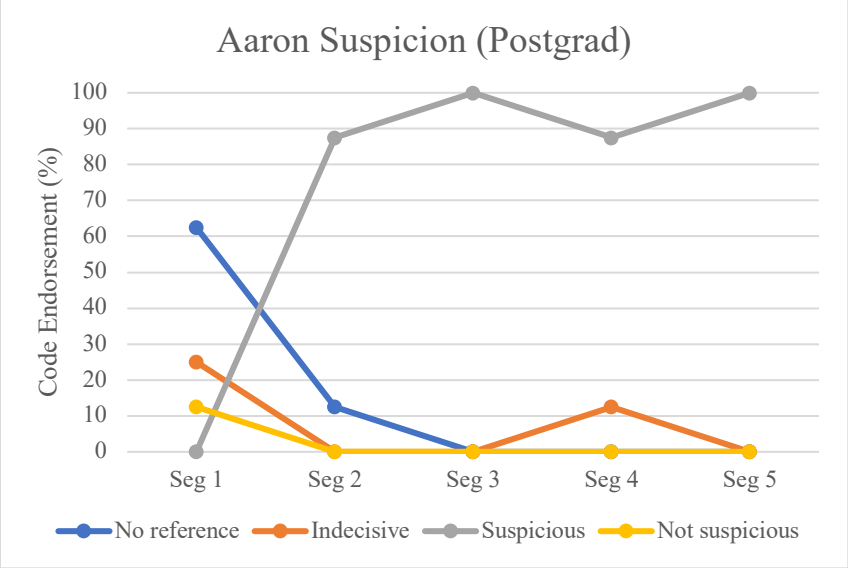
(3) Not suspicious
- The participant references suspicion but states they believe the situation is not suspicious, or that they have no reason to be suspicious at that time
    - "I have no reason to accuse him of anything yet"
    - "I don't think I'm doing anything wrong"
    - "I see no reason to be suspicious at this point"
    - "I see no reason to get upset because the girl looks different. Who am I to judge?"

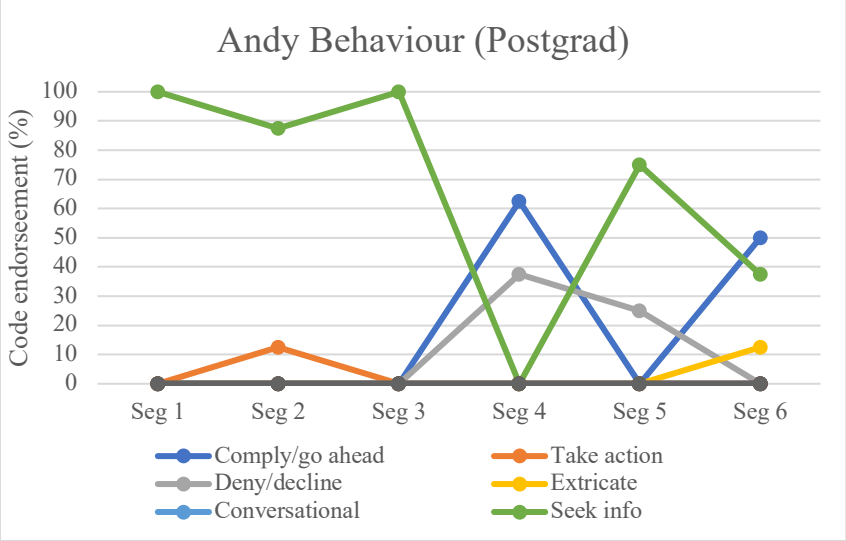
If participant makes reference to suspicion in the 'behaviour' section (i.e. *What would you do?),* still count this and code under the 'explanation' section.

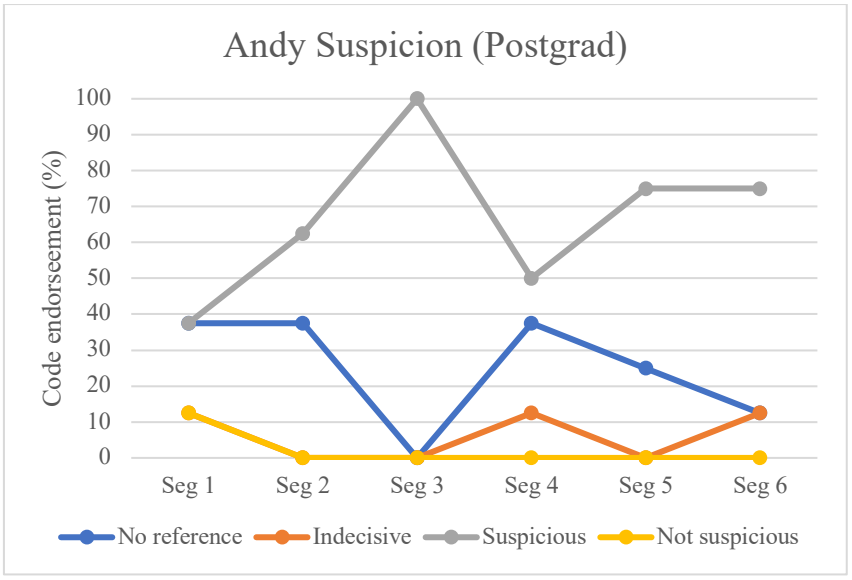
Appendix C

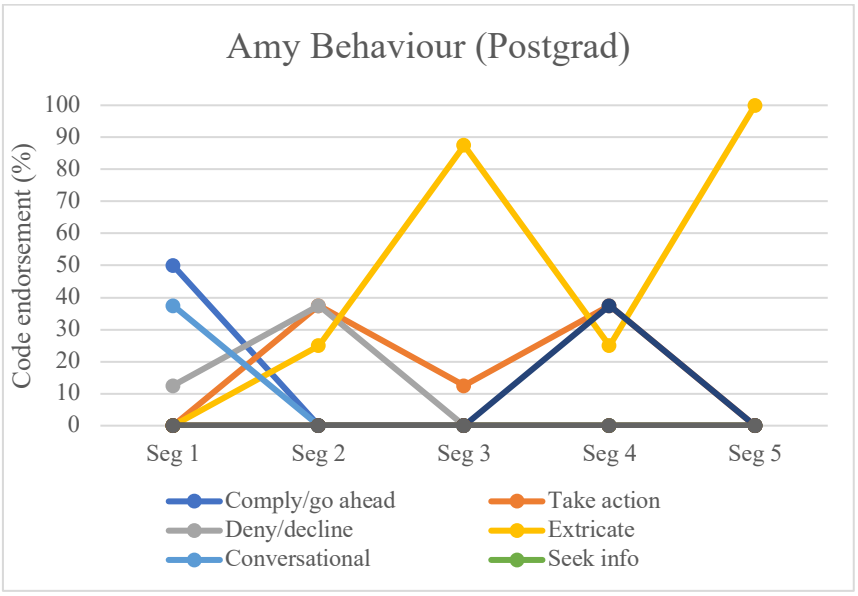
Postgrad psychology students (N=8)

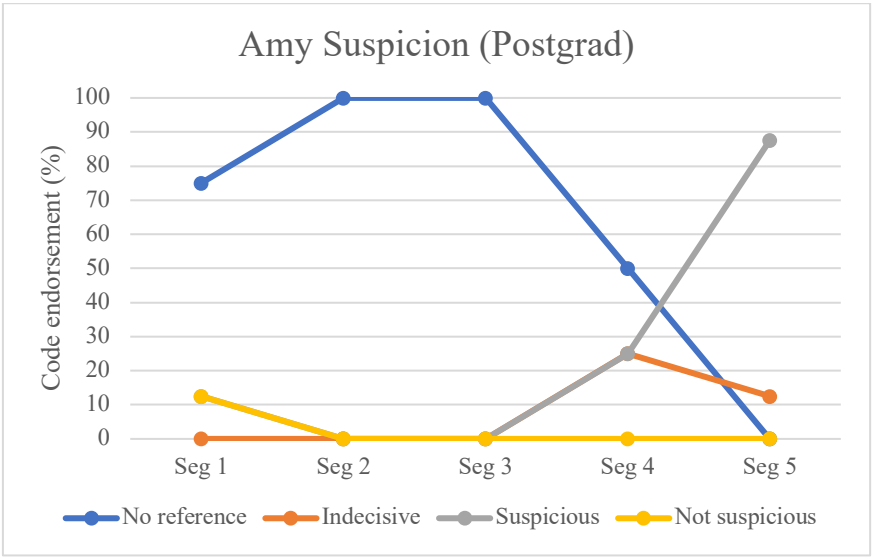


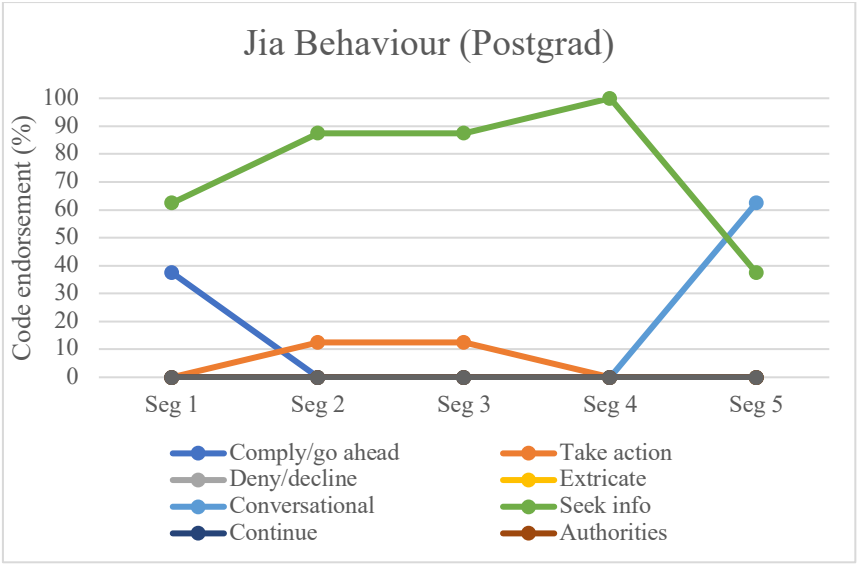


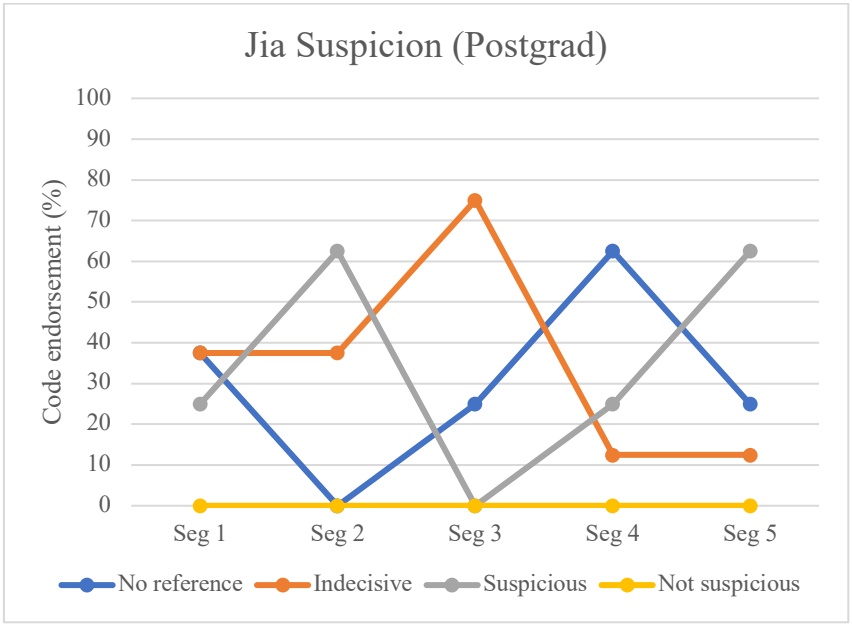


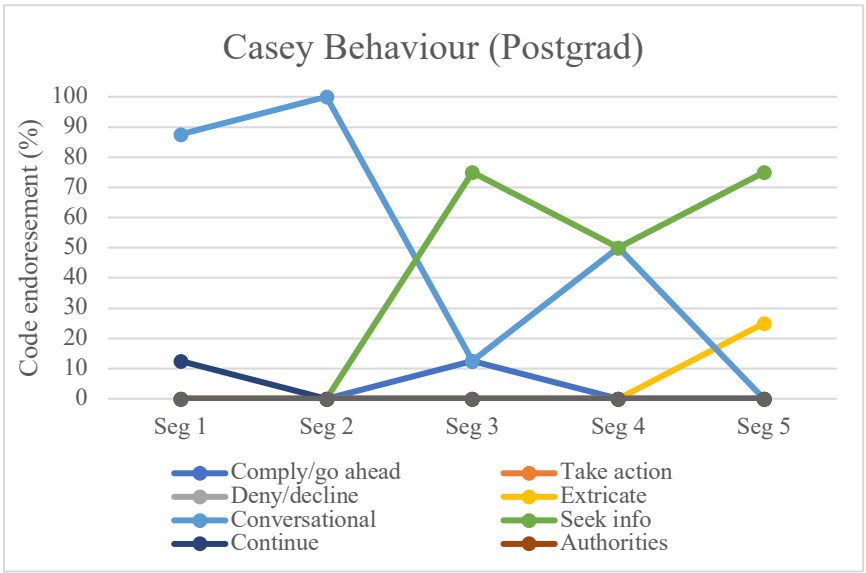

Andy Suspicion (Postgrad)
Code endorseement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Seg 6
No reference
Indecisive
Suspicious
Not suspicious

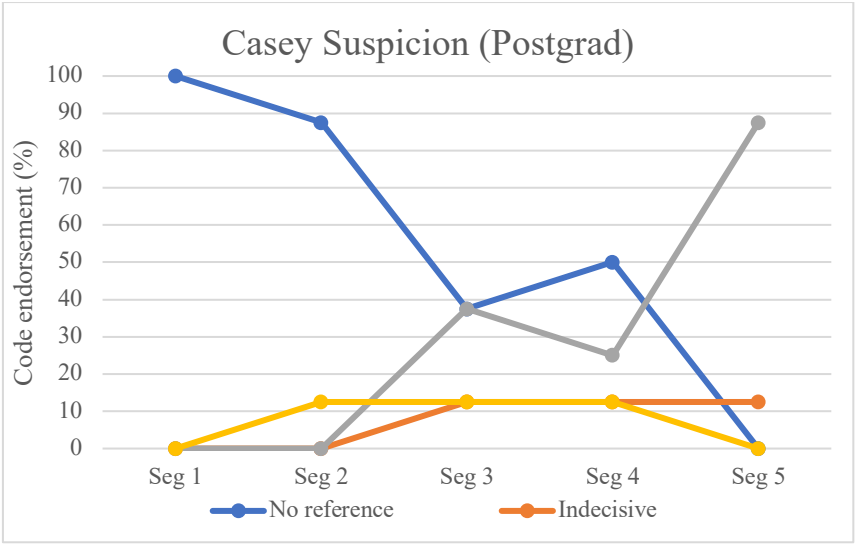

Amy Behaviour (Postgrad)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Comply/go ahead
Deny/decline
Conversational
Take action
Extricate
Seek info

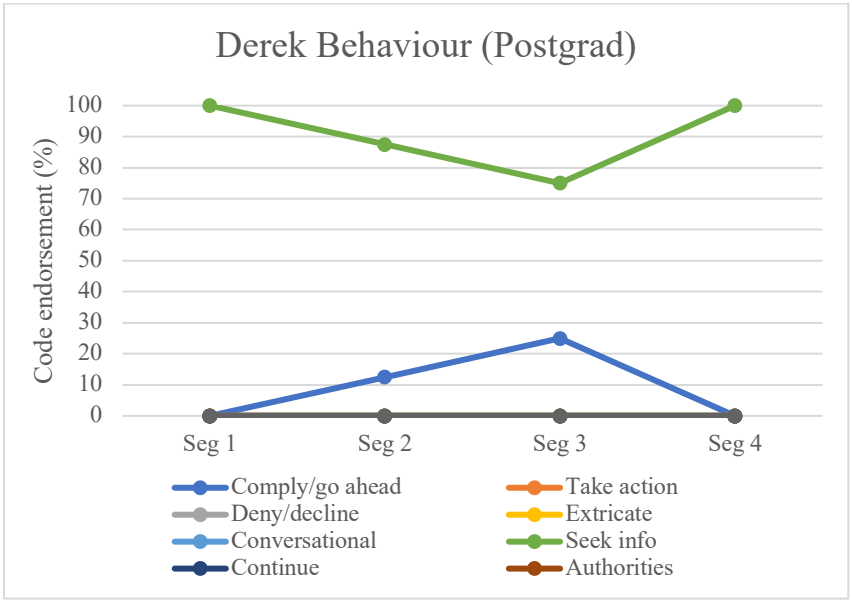

Amy Suspicion (Postgrad)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
No reference
Indecisive
Suspicious
Not suspicious

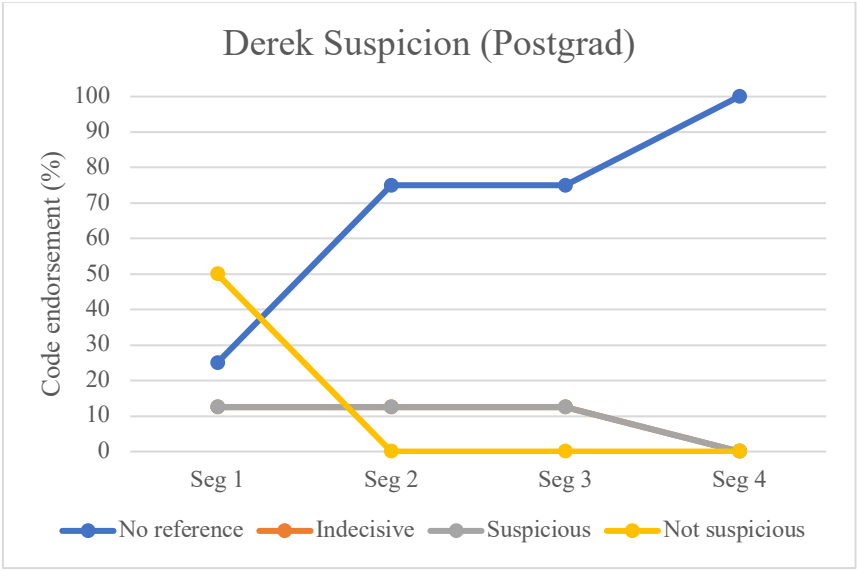

Jia Behaviour (Postgrad)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Comply/go ahead
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue
Authorities


Jia Suspicion (Postgrad)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
No reference
Indecisive
Suspicious
Not suspicious


Casey Behaviour (Postgrad)
Code endoresement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Comply/go ahead
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue
Authorities


Casey Suspicion (Postgrad)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
No reference
Indecisive


Derek Behaviour (Postgrad)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Comply/go ahead
Deny/decline
Conversational
Continue
Take action
Extricate
Seek info
Authorities


Derek Suspicion (Postgrad)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
No reference
Indecisive
Suspicious
Not suspicious

Appendix D

Legal Sample (N=4)

Aaron	Behaviour	%	Suspicion	%
Segment 1	Conversational	75	No reference	100
	Take action	25		
Segment 2	Accept offer	50	No reference	50
	Decline offer	25	Suspicious	50
	Seek info	25		
Segment 3	Decline offer	50	Indecisive reference	50
	Seek info	50	Suspicious	50
Segment 4	Extricate	50	Suspicious	100
	Seek info	25		
	Involve authorities	25		
Segment 5	Extricate	25	Suspicious	75
	Seek info	25	Not suspicious	25
	Continue	25		
	Involve authorities	25		

Andy	Behaviour	%	Suspicion	%
Segment 1	Accept offer	25	No reference	75
	Seek info	75	Indecisive reference	25
Segment 2	Accept offer	25	Indecisive reference	50
	Decline offer	25	Suspicious	50
	Seek info	50		
Segment 3	Extricate	50	Indecisive reference	25
	Seek info	50	Suspicious	75
Segment 4	Comply	25	No reference	50
	Extricate	25	Suspicious	50
	Seek info	50		
Segment 5	Seek info	100	Indecisive reference	50
			Suspicious	50
Segment 6	Comply	25	No reference	25
	Seek info	75	Suspicious	75

Amy	Behaviour	%	Suspicion	%
Segment 1	Conversational	100	No reference	25
			Not suspicious	75
Segment 2	Take action	25	No reference	100
	Decline	25		
	Extricate	50		
Segment 3	Comply	25	No reference	100
	Extricate	75		
Segment 4	Take action	100	No reference	50
			Indecisive reference	50

Segment 5	Take action	75	Indecisive reference	25
	Extricate	25	Suspicious	75

Jia	Behaviour	%	Suspicion	%
Segment 1	Accept offer	75	No reference	75
	Seek info	25	Indecisive reference	25
Segment 2	Go ahead	75	No reference	25
	Seek info	25	Indecisive reference	75
Segment 3	Go ahead	50	No reference	25
	Seek info	50	Indecisive reference	50
			Suspicious	25
Segment 4	Take action	25	No reference	50
	Seek info	75	Suspicious	50
Segment 5	Seek info	50	Suspicious	100
	Involve authorities	50		

Casey	Behaviour		Suspicion	
Segment 1	Conversational	100%	No reference	100%
Segment 2	Conversational	100%	No reference	100%
Segment 3	Go ahead	25%	No reference	100%
	Seek info	75%		
Segment 4	Take action	25%	No reference	50%
	Conversational	25%	Indecisive reference	50%
	Seek info	25%		
	Non-behavioural	25%		
Segment 5	Extricate	75%	Suspicious	100%
	Seek info	25%		

Derek	Behaviour		Suspicion	
Segment 1	Seek info	100%	No reference	25%
			Indecisive reference	75%
Segment 2	Seek info	100%	Indecisive reference	100%
Segment 3	Go ahead	50%	No reference	25%
	Seek info	50%	Indecisive reference	25%
			Suspicious	50%
Segment 4	Seek info	100%	No reference	75%
			Indecisive reference	25%

Appendix E

Round 1: *Complete agreement (1 code the same & 1 different = incorrect)*

Scenario	Behaviour	Suspicion
Aaron	.74	.70
Andy	.71	.73
Amy	.63	.56
Jia	.64	.69
Casey	.58	.65
Derek	.57	.66

Round 1: *partial agreement (at least 1 code the same = correct)*

Scenario	Behaviour	Suspicion
Aaron	.83	.70
Andy	.79	.73
Amy	.83	.58
Jia	.78	.71
Casey	.72	.65
Derek	.60	.66

Round 2

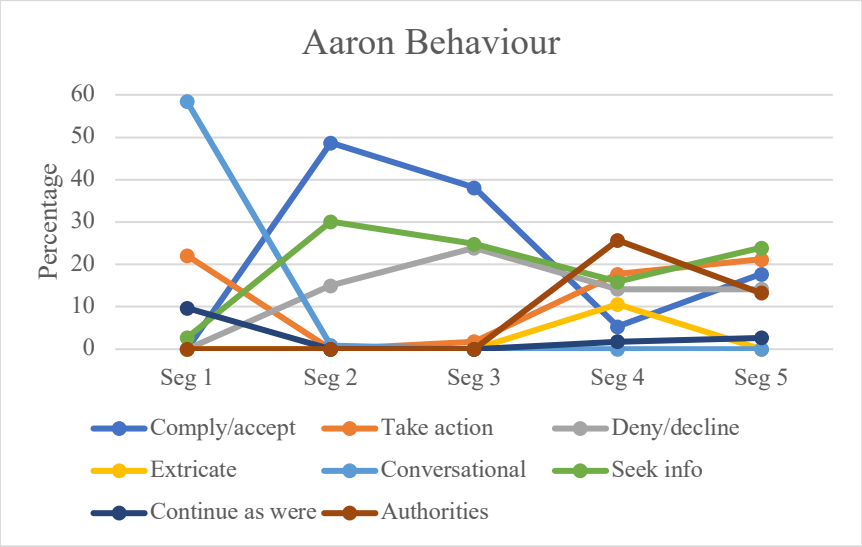
Round 2: *Complete agreement (1 code the same & 1 different = incorrect)*

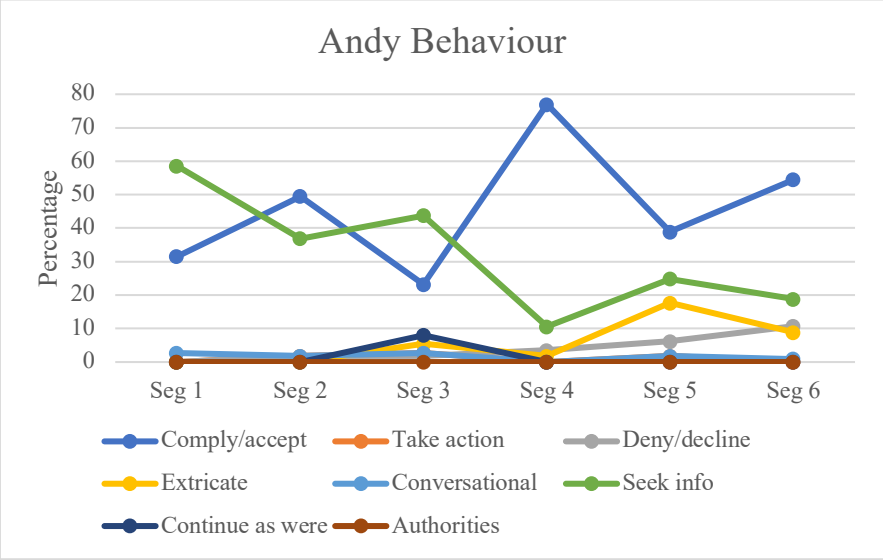
Scenario	Behaviour	Suspicion
Aaron	.74	.74
Andy	.59	.79
Amy	.58	.82
Jia	.75	.73
Casey	.55	.81
Derek	.80	.60

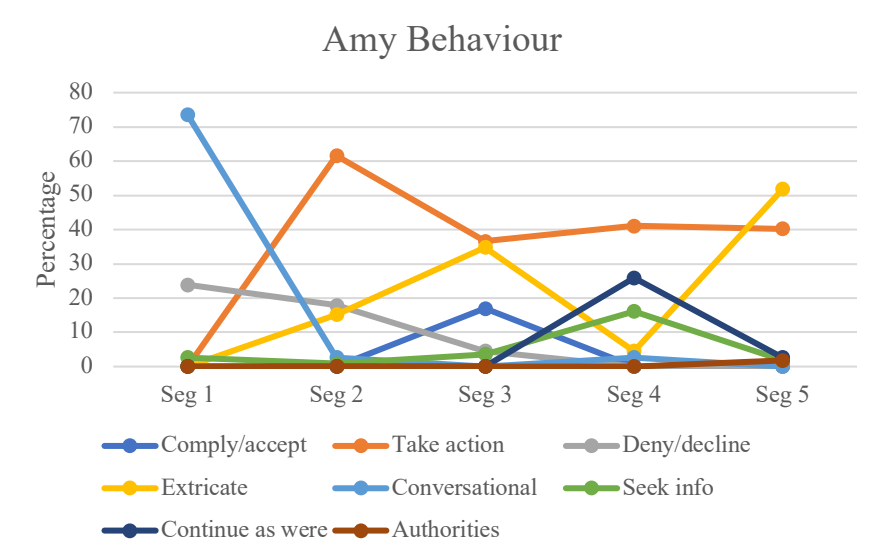
Round 2: *Partial agreement (At least 1 code the same = right)*

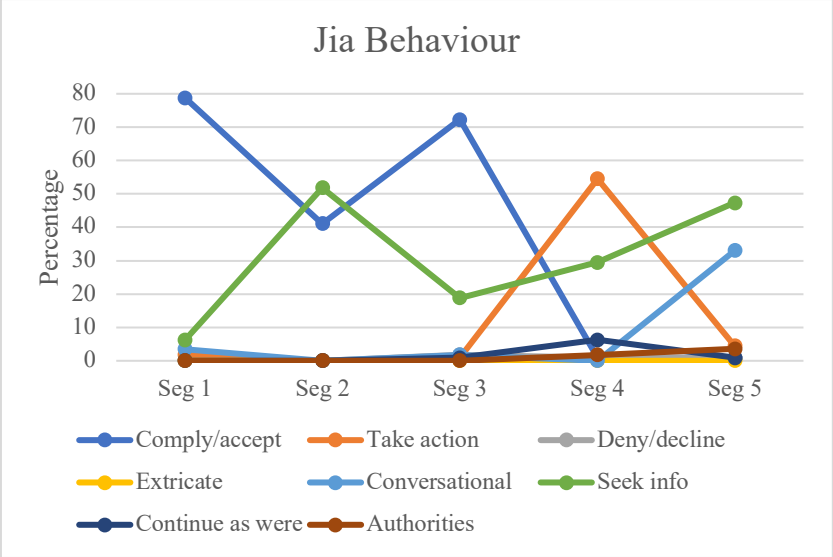
Scenario	Behaviour	Suspicion
Aaron	.89	.74
Andy	.82	.79
Amy	.88	.82
Jia	.82	.75
Casey	.82	.81
Derek	.95	.60

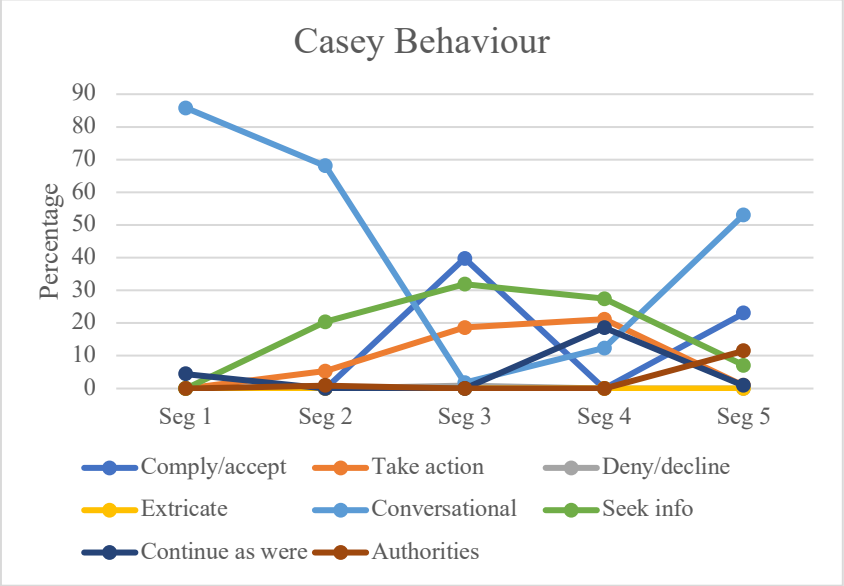
Appendix F

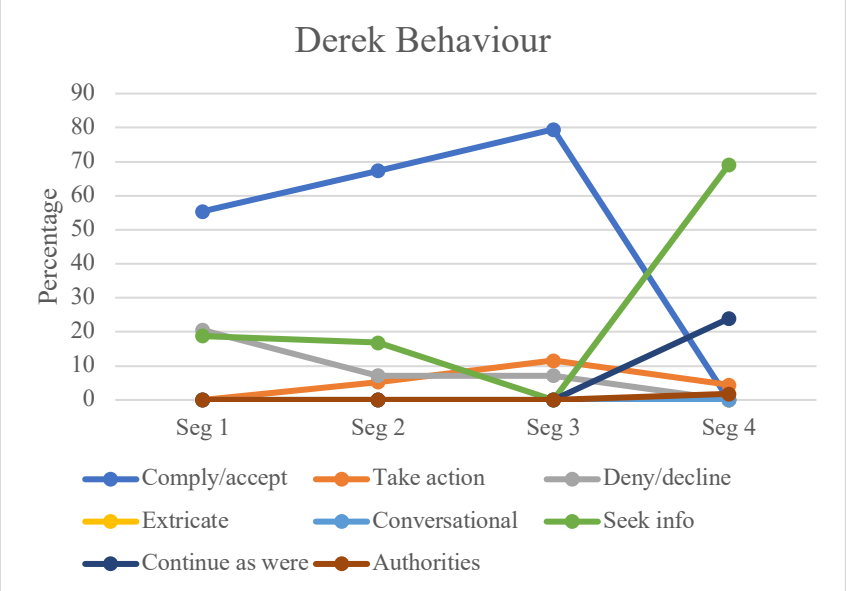

Aaron Behaviour
Percentage
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Comply/accept
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue as were
Authorities


Andy Behaviour
Percentage
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Seg 6
Comply/accept
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue as were
Authorities


Amy Behaviour
Percentage
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Comply/accept
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue as were
Authorities


Jia Behaviour
Percentage
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Comply/accept
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue as were
Authorities


Casey Behaviour
Percentage
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Comply/accept
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue as were
Authorities


Derek Behaviour
Percentage
90
80
70
60
50
40
30
20
10
0
Seg 1
Seg 2
Seg 3
Seg 4
Comply/accept
Take action
Deny/decline
Extricate
Conversational
Seek info
Continue as were
Authorities

Appendix G

Table 1

One Sample Binomial Test

Aaron Behaviour	N	Observed P	*p**	Cohen's *g**
Segment 1				
Don't extricate[1] (seek info)[2]	77[1] (3)[2]	.96[1] (.04)[2]	<.001	.46
Don't extricate (extricate)	77 (25)	.75 (.25)	<.001	.25
Seek info (extricate)	3 (25)	.89 (.11)	<.001	.39
Segment 2				
Don't extricate (seek info)	57 (34)	.63 (.37)	.062	.13
Don't extricate (extricate)	57 (17)	.77 (.23)	<.001	.27
Seek info (extricate)	34 (17)	.67 (.33)	.072	.17
Segment 3				
Don't extricate (seek info)	48 (28)	.63 (.37)	.086	.13
Don't extricate (extricate)	48 (29)	.62 (.38)	.119	.12
Seek info (extricate)	28 (29)	.49 (.51)	1.00	.01
Segment 4				
Don't extricate (seek info)	8 (18)	.31 (.69)	.227	.19
Don't extricate (extricate)	8 (82)	.09 (.91)	<.001	.41
Seek info (extricate)	18 (82)	.18 (.82)	<.001	.32
Segment 5				
Don't extricate (seek info)	24 (27)	.47 (.53)	1.00	.03
Don't extricate (extricate)	24 (55)	.30 (.70)	.002	.20
Seek info (extricate)	27 (55)	.33 (.67)	.008	.17

***Test proportion = 0.5**

***Bonferroni correction**

Table 2

One Sample Binomial Test

Aaron Suspicion	N	Observed P	*p**	Cohen's *g**
Segment 1				
No reference (indecisive)	109 (3)	.97 (.03)	<.001	.47
No reference (suspicious)	109 (0)	1.00 (0.00)	<.001	.50
No reference (not suspicious)	109 (1)	.99 (.01)	<.001	.49
Indecisive (suspicious)	3 (0)	1.00 (0.00)	1.00	.50
Indecisive (not suspicious)	3 (1)	.75 (.25)	1.00	.25
Suspicious (not suspicious)	0 (1)	-	-	-
Segment 2				
No reference (indecisive)	59 (8)	.88 (.12)	<.001	.38
No reference (suspicious)	59 (42)	.58 (.42)	.666	.08
No reference (not suspicious)	59 (3)	.95 (.05)	<.001	.45
Indecisive (suspicious)	8 (42)	.16 (.84)	<.001	.34
Indecisive (not suspicious)	8 (3)	.73 (.27)	1.00	.23
Suspicious (not suspicious)	42 (3)	.93 (.07)	<.001	.43
Segment 3				
No reference (indecisive)	37 (5)	.88 (.12)	<.001	.38
No reference (suspicious)	37 (69)	.35 (.65)	.015	.15
No reference (not suspicious)	37 (2)	.95 (.05)	<.001	.45
Indecisive (suspicious)	5 (69)	.07 (.93)	<.001	.43
Indecisive (not suspicious)	5 (2)	.71 (.29)	1.00	.21
Suspicious (not suspicious)	69 (2)	.97 (.03)	<.001	.47
Segment 4				
No reference (indecisive)	8 (11)	.42 (.58)	1.00	.08
No reference (suspicious)	8 (94)	.08 (.92)	<.001	.42
No reference (not suspicious)	8 (0)	1.00 (0.00)	.047	.50
Indecisive (suspicious)	11 (94)	.10 (.90)	<.001	.40
Indecisive (not suspicious)	11 (0)	1.00 (0.00)	.006	.50
Suspicious (not suspicious)	94 (0)	1.00 (0.00)	<.001	.50
Segment 5				
No reference (indecisive)	9 (15)	.63 (.38)	1.00	.13
No reference (suspicious)	9 (82)	.10 (.90)	<.001	.40
No reference (not suspicious)	9 (7)	.56 (.44)	1.00	.06
Indecisive (suspicious)	15 (82)	.15 (.85)	<.001	.35
Indecisive (not suspicious)	15 (7)	.68 (.32)	.803	.18
Suspicious (not suspicious)	82 (7)	.92 (.08)	<.001	.42

Table 3

One Sample Binomial Test

Andy Behaviour	N	Observed Prop	*p**	Cohen's *g**
Segment 1				
Don't extricate[1] (seek info)[2]	41[1] (65)[2]	.39[1] (.61)[2]	.075	.11
Don't extricate (extricate)	41 (1)	.98 (.02)	<.001	.48
Seek info (extricate)	65 (1)	.98 (.02)	<.001	.48
Segment 2				
Don't extricate (seek info)	64 (41)	.61 (.39)	.094	.11
Don't extricate (extricate)	64 (4)	.94 (.06)	<.001	.44
Seek info (extricate)	41 (4)	.91 (.09)	<.001	.41
Segment 3				
Don't extricate (seek info)	44 (49)	.47 (.53)	1.00	.03
Don't extricate (extricate)	44 (11)	.80 (.20)	<.001	.30
Seek info (extricate)	49 (11)	.82 (.18)	<.001	.32
Segment 4				
Don't extricate (seek info)	93 (12)	.89 (.11)	<.001	.39
Don't extricate (extricate)	93 (6)	.94 (.06)	<.001	.44
Seek info (extricate)	12 (6)	.67 (.33)	.714	.17
Segment 5				
Don't extricate (seek info)	50 (28)	.64 (.36)	.051	.14
Don't extricate (extricate)	50 (29)	.63 (.37)	.071	.13
Seek info (extricate)	28 (29)	.49 (.51)	1.00	.01
Segment 6				
Don't extricate (seek info)	66 (21)	.76 (.24)	<.001	.26
Don't extricate (extricate)	66 (22)	.75 (.25)	<.001	.25
Seek info (extricate)	21 (22)	.49 (.51)	1.00	.01

***Test proportion = 0.5**

***Bonferroni correction**

Table 4

One Sample Binomial Test

Andy Suspicion	N	Observed P	*p**	Cohen's *g**
Segment 1				
No reference (indecisive)	87 (14)	.96 (.14)	<.001	.46
No reference (suspicious)	87 (8)	.92 (.08)	<.001	.42
No reference (not suspicious)	87 (2)	.98 (.02)	<.001	.48
Indecisive (suspicious)	14 (8)	.64 (.36)	1.00	.14
Indecisive (not suspicious)	14 (2)	.88 (.13)	.025	.38
Suspicious (not suspicious)	8 (2)	.80 (.20)	.656	.30
Segment 2				
No reference (indecisive)	71 (7)	.91 (.09)	<.001	.41
No reference (suspicious)	71 (26)	.73 (.27)	<.001	.23
No reference (not suspicious)	71 (7)	.91 (.09)	<.001	.41
Indecisive (suspicious)	7 (26)	.21 (.79)	.008	.29
Indecisive (not suspicious)	7 (7)	.50 (.50)	1.00	0
Suspicious (not suspicious)	26 (7)	.79 (.21)	.008	.29
Segment 3				
No reference (indecisive)	60 (8)	.88 (.12)	<.001	.38
No reference (suspicious)	60 (39)	.61 (.39)	.263	.11
No reference (not suspicious)	60 (5)	.92 (.08)	<.001	.42
Indecisive (suspicious)	8 (39)	.17 (.83)	<.001	.33
Indecisive (not suspicious)	8 (5)	.62 (.38)	1.00	.12
Suspicious (not suspicious)	39 (5)	.89 (.11)	<.001	.39
Segment 4				
No reference (indecisive)	71 (7)	.91 (.09)	<.001	.41
No reference (suspicious)	71 (25)	.74 (.26)	<.001	.24
No reference (not suspicious)	71 (10)	.88 (.12)	<.001	.38
Indecisive (suspicious)	7 (25)	.22 (.78)	.013	.28
Indecisive (not suspicious)	10 (7)	.41 (.59)	1.00	.09
Suspicious (not suspicious)	25 (10)	.71 (.29)	.100	.21
Segment 5				
No reference (indecisive)	32 (10)	.76 (.24)	.006	.26
No reference (suspicious)	32 (60)	.35 (.65)	.028	.15
No reference (not suspicious)	32 (11)	.74 (.26)	.011	.24
Indecisive (suspicious)	10 (60)	.14 (.86)	<.001	.36
Indecisive (not suspicious)	10 (11)	.48 (.52)	1.00	.02
Suspicious (not suspicious)	11 (60)	.85 (.15)	<.001	.35
Segment 6				
No reference (indecisive)	47 (9)	.84 (.16)	<.001	.34
No reference (suspicious)	47 (54)	.47 (.53)	1.00	.03
No reference (not suspicious)	47 (2)	.96 (.04)	<.001	.46
Indecisive (suspicious)	9 (54)	.14 (.86)	<.001	.36
Indecisive (not suspicious)	9 (2)	.82 (.18)	.393	.32
Suspicious (not suspicious)	52 (2)	.96 (.04)	<.001	.46

Table 5

One Sample Binomial Test

Amy Behaviour	N	Observed P	*p**	Cohen's *g*
Segment 1				
Don't extricate (seek info)	83 (3)	.97 (.03)	<.001	.47
Don't extricate (extricate)	83 (27)	.75 (.25)	<.001	.25
Seek info (extricate)	3 (27)	.10 (.90)	<.001	.40
Segment 2				
Don't extricate (seek info)	3 (1)	.75 (.25)	1.00	.25
Don't extricate (extricate)	3 (106)	.03 (.97)	<.001	.47
Seek info (extricate)	106 (1)	.01 (.99)	<.001	.49
Segment 3				
Don't extricate (seek info)	19 (4)	.83 (.17)	.008	.33
Don't extricate (extricate)	19 (85)	.18 (.82)	<.001	.32
Seek info (extricate)	4 (85)	.04 (.96)	<.001	.46
Segment 4				
Don't extricate (seek info)	32 (18)	.64 (.36)	.195	.14
Don't extricate (extricate)	32 (51)	.39 (.61)	.143	.11
Seek info (extricate)	51 (18)	.26 (.74)	<.001	.24
Segment 5				
Don't extricate (seek info)	3 (2)	.60 (.40)	1.00	.10
Don't extricate (extricate)	3 (105)	.03 (.97)	<.001	.47
Seek info (extricate)	2 (105)	.02 (.98)	<.001	.48

***Test proportion = 0.5**

***Bonferroni correction**

Table 6

One Sample Binomial Test

Amy Suspicion	N	Observed P	*p**	Cohen's *g**
Segment 1				
No reference (indecisive)	95	.72 (.28)	<.001	.22
No reference (suspicious)	76	.89 (.11)	<.001	.39
No reference (not suspicious)	78	.87 (.13)	<.001	.37
Indecisive (suspicious)	36	.77 (.23)	.011	.27
Indecisive (not suspicious)	37	.73 (.27)	.046	.23
Suspicious (not suspicious)	18	.44 (.56)	1.00	.06
Segment 2				
No reference (indecisive)	97	.92 (.08)	<.001	.42
No reference (suspicious)	103	.82 (.18)	<.001	.32
No reference (not suspicious)	84	1.00 (0.00)	<.001	.50
Indecisive (suspicious)	26	.27 (.73)	.174	.23
Indecisive (not suspicious)	7	1.00 (0.00)	.094	.50
Suspicious (not suspicious)	19	1.00 (0.00)	<.001	.50
Segment 3				
No reference (indecisive)	83	.83 (.17)	<.001	.33
No reference (suspicious)	93	.74 (.26)	<.001	.24
No reference (not suspicious)	73	.95 (.05)	<.001	.45
Indecisive (suspicious)	38	.37 (.63)	.860	.13
Indecisive (not suspicious)	18	.78 (.22)	.185	.28
Suspicious (not suspicious)	28	.86 (.14)	.001	.36
Segment 4				
No reference (indecisive)	58	.84 (.16)	<.001	.34
No reference (suspicious)	103	.48 (.52)	1.00	.02
No reference (not suspicious)	49	1.00 (0.00)	<.001	.50
Indecisive (suspicious)	63	.14 (.86)	<.001	.36
Indecisive (not suspicious)	9	1.00 (0.00)	.023	.50
Suspicious (not suspicious)	54	1.00 (0.00)	<.001	.50
Segment 5				
No reference (indecisive)	11	.36 (.64)	1.00	.14
No reference (suspicious)	105	.04 (.96)	<.001	.46
No reference (not suspicious)	4	1.00 (0.00)	.750	.50
Indecisive (suspicious)	108	.06 (.94)	<.001	.44
Indecisive (not suspicious)	7	1.00 (0.00)	.094	.50
Suspicious (not suspicious)	101	1.00 (0.00)	<.001	.50

Table 7

One Sample Binomial Test

Jia Behaviour	N	Observed P	*p**	Cohen's *g*
Segment 1				
Don't extricate (seek info)	94 (7)	.93 (.07)	<.001	.43
Don't extricate (extricate)	94 (2)	.98 (.02)	<.001	.48
Seek info (extricate)	7 (2)	.78 (.22)	.539	.28
Segment 2				
Don't extricate (seek info)	50 (58)	.46 (.54)	1.00	.04
Don't extricate (extricate)	50 (0)	1.00 (0.00)	<.001	.50
Seek info (extricate)	58 (0)	1.00 (0.00)	<.001	.50
Segment 3				
Don't extricate (seek info)	84 (21)	.80 (.20)	<.001	.30
Don't extricate (extricate)	84 (3)	.97 (.03)	<.001	.47
Seek info (extricate)	21 (3)	.88 (.13)	.001	.38
Segment 4				
Don't extricate (seek info)	8 (33)	.20 (.80)	<.001	.30
Don't extricate (extricate)	8 (64)	.11 (.89)	<.001	.39
Seek info (extricate)	33 (64)	.34 (.66)	.006	.16
Segment 5				
Don't extricate (seek info)	39 (53)	.42 (.58)	.525	.08
Don't extricate (extricate)	39 (10)	.80 (.20)	<.001	.30
Seek info (extricate)	53 (10)	.84 (.16)	<.001	.34

Table 8

One Sample Binomial Test

Jia Suspicion	N	Observed P	*p**	Cohen's *g**
Segment 1				
No reference (indecisive)	105	.96 (.04)	<.001	.46
No reference (suspicious)	103	.89 (.02)	<.001	.39
No reference (not suspicious)	102	.99 (.01)	<.001	.49
Indecisive (suspicious)	6	.67 (.33)	1.00	.17
Indecisive (not suspicious)	5	.80 (.20)	1.00	.30
Suspicious (not suspicious)	3	.67 (.33)	1.00	.17
Segment 2				
No reference (indecisive)	97	.91 (.09)	<.001	.41
No reference (suspicious)	99	.89 (.11)	<.001	.39
No reference (not suspicious)	89	.99 (.01)	<.001	.49
Indecisive (suspicious)	20	.45 (.55)	1.00	.05
Indecisive (not suspicious)	10	.90 (.10)	.129	.40
Suspicious (not suspicious)	12	.92 (.08)	.038	.42
Segment 3				
No reference (indecisive)	96	.93 (.07)	<.001	.43
No reference (suspicious)	103	.86 (.14)	<.001	.36
No reference (not suspicious)	91	.98 (.02)	<.001	.48
Indecisive (suspicious)	21	.33 (.67)	1.00	.17
Indecisive (not suspicious)	9	.78 (.22)	1.00	.28
Suspicious (not suspicious)	16	.88 (.13)	.025	.38
Segment 4				
No reference (indecisive)	90	.98 (.02)	<.001	.48
No reference (suspicious)	108	.81 (.19)	<.001	.31
No reference (not suspicious)	89	.99 (.01)	<.001	.49
Indecisive (suspicious)	22	.09 (.91)	.001	.41
Indecisive (not suspicious)	3	.67 (.33)	1.00	.17
Suspicious (not suspicious)	21	.95 (.05)	<.001	.45
Segment 5				
No reference (indecisive)	46	.72 (.28)	.027	.22
No reference (suspicious)	98	.34 (.66)	.010	.16
No reference (not suspicious)	34	.97 (.03)	<.001	.47
Indecisive (suspicious)	78	.17 (.83)	<.001	.33
Indecisive (not suspicious)	14	.93 (.07)	.011	.43
Suspicious (not suspicious)	66	.98 (.02)	<.001	.48

Table 9

One Sample Binomial Test

Casey Behaviour	N	Observed P	*p**	Cohen's *g*
Segment 1				
Don't extricate (seek info)	102 (0)	1.00 (0.00)	<.001	.50
Don't extricate (extricate)	102 (0)	1.00 (0.00)	<.001	.50
Seek info (extricate)	0 (0)	-	-	-
Segment 2				
Don't extricate (seek info)	77 (23)	.77 (.23)	<.001	.27
Don't extricate (extricate)	77 (7)	.92 (.08)	<.001	.42
Seek info (extricate)	23 (7)	.77 (.23)	.016	.27
Segment 3				
Don't extricate (seek info)	50 (36)	.58 (.42)	.482	.08
Don't extricate (extricate)	50 (22)	.69 (.31)	.004	.19
Seek info (extricate)	36 (22)	.62 (.38)	.261	.12
Segment 4				
Don't extricate (seek info)	35 (31)	.53 (.47)	1.00	.03
Don't extricate (extricate)	35 (24)	.59 (.41)	.578	.09
Seek info (extricate)	31 (24)	.56 (.44)	1.00	.06
Segment 5				
Don't extricate (seek info)	87 (8)	.92 (.08)	<.001	.42
Don't extricate (extricate)	87 (14)	.86 (.14)	<.001	.36
Seek info (extricate)	8 (14)	.36 (.64)	.859	.14

Table 10

One Sample Binomial Test

Casey suspicion	N	Observed P	*p**	Cohen's *g**
Segment 1				
No reference (indecisive)	111	.99 (.01)	<.001	.44
No reference (suspicious)	110	1.00 (.00)	<.001	.50
No reference (not suspicious)	110	1.00 (.00)	<.001	.50
Indecisive (suspicious)	0	-	-	-
Indecisive (not suspicious)	0	-	-	-
Suspicious (not suspicious)	0	-	-	-
Segment 2				
No reference (indecisive)	94	.95 (.05)	<.001	.45
No reference (suspicious)	107	.83 (.17)	<.001	.33
No reference (not suspicious)	90	.99 (.01)	<.001	.49
Indecisive (suspicious)	23	.22 (.78)	.064	.28
Indecisive (not suspicious)	6	.83 (.17)	1.00	.33
Suspicious (not suspicious)	19	.95 (.05)	<.001	.45
Segment 3				
No reference (indecisive)	75	.92 (.08)	<.001	.42
No reference (suspicious)	100	.69 (.31)	.001	.19
No reference (not suspicious)	74	.93 (.07)	<.001	.43
Indecisive (suspicious)	37	.16 (.84)	<.001	.34
Indecisive (not suspicious)	11	.55 (.45)	1.00	.05
Suspicious (not suspicious)	36	.86 (.14)	<.001	.36
Segment 4				
No reference (indecisive)	51	.88 (.12)	<.001	.38
No reference (suspicious)	102	.44 (.56)	1.00	.06
No reference (not suspicious)	.49	.92 (.08)	<.001	.42
Indecisive (suspicious)	63	.10 (.90)	<.001	.40
Indecisive (not suspicious)	10	.60 (.40)	1.00	.10
Suspicious (not suspicious)	61	.93 (.07)	<.001	.43
Segment 5				
No reference (indecisive)	40	.07 (.93)	<.001	.43
No reference (suspicious)	75	.04 (.96)	<.001	.46
No reference (not suspicious)	3	1.00 (0.00)	1.00	.50
Indecisive (suspicious)	109	.34 (.66)	.006	.16
Indecisive (not suspicious)	37	1.00 (0.00)	<.001	.50
Suspicious (not suspicious)	72	1.00 (0.00)	<.001	.50

Table 11

One Sample Binomial Test				
Derek behaviour	N	Observed P	*p**	Cohen's *g*
Segment 1				
Don't extricate (seek info)	68 (21)	.76 (24)	<.001	.26
Don't extricate (extricate)	68 (23)	.75 (.25)	<.001	.25
Seek info (extricate)	21 (23)	.48 (.52)	1.00	.02
Segment 2				
Don't extricate (seek info)	80 (19)	.81 (.19)	<.001	.31
Don't extricate (extricate)	80 (14)	.85 (.15)	<.001	.35
Seek info (extricate)	19 (14)	.58 (.42)	1.00	.08
Segment 3				
Don't extricate (seek info)	90 (8)	.92 (.08)	<.001	.42
Don't extricate (extricate)	90 (13)	.87 (.13)	<.001	.37
Seek info (extricate)	8 (13)	.38 (.62)	1.00	.12
Segment 4				
Don't extricate (seek info)	27 (77)	.26 (.74)	<.001	.24
Don't extricate (extricate)	27 (8)	.77 (.23)	.006	.27
Seek info (extricate)	77 (8)	.91 (.09)	<.001	.41

Table 12

One Sample Binomial Test

Derek suspicion	N	Observed P	*p**	Cohen's *g**
Segment 1				
No reference (indecisive)	84	.89 (.11)	<.001	.39
No reference (suspicious)	96	.78 (.22)	<.001	.28
No reference (not suspicious)	82	.91 (.09)	<.001	.41
Indecisive (suspicious)	30	.30 (.70)	.257	.20
Indecisive (not suspicious)	16	.56 (.44)	1.00	.05
Suspicious (not suspicious)	28	.75 (.25)	.075	.25
Segment 2				
No reference (indecisive)	92	.96 (.04)	<.001	.46
No reference (suspicious)	105	.84 (.16)	<.001	.34
No reference (not suspicious)	92	.96 (.04)	<.001	.46
Indecisive (suspicious)	21	.19 (.81)	.043	.31
Indecisive (not suspicious)	8	.50 (.50)	1.00	0
Suspicious (not suspicious)	21	.81 (.19)	.043	.31
Segment 3				
No reference (indecisive)	87	.86 (.14)	<.001	.36
No reference (suspicious)	90	.83 (.17)	<.001	.33
No reference (not suspicious)	85	.88 (.12)	<.001	.38
Indecisive (suspicious)	27	.44 (.56)	1.00	.06
Indecisive (not suspicious)	22	.55 (.45)	1.00	.05
Suspicious (not suspicious)	25	.60 (.40)	1.00	.10
Segment 4				
No reference (indecisive)	99	1.00 (0.00)	<.001	.50
No reference (suspicious)	113	.88 (.12)	<.001	.38
No reference (not suspicious)	99	1.00 (0.00)	<.001	.50
Indecisive (suspicious)	14	1.00 (0.00)	.001	.50
Indecisive (not suspicious)	0	-	-	-
Suspicious (not suspicious)	14	1.00 (0.00)	.001	.50

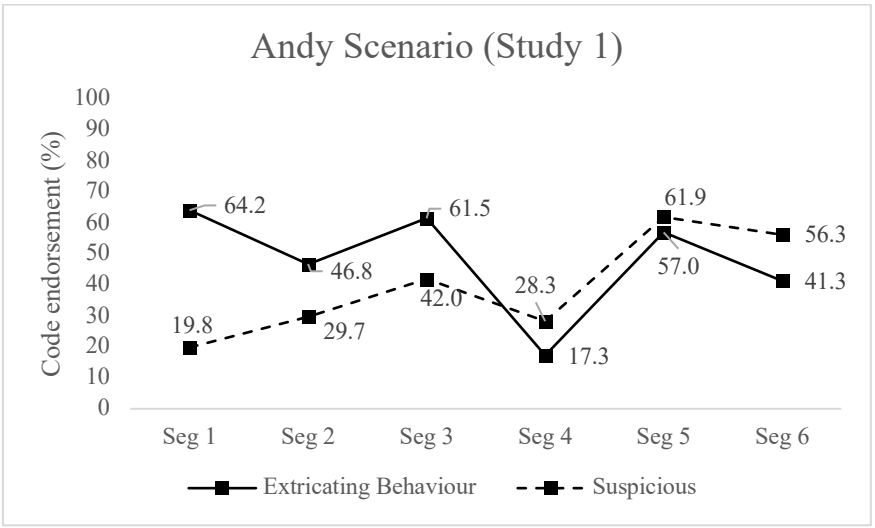
Appendix H

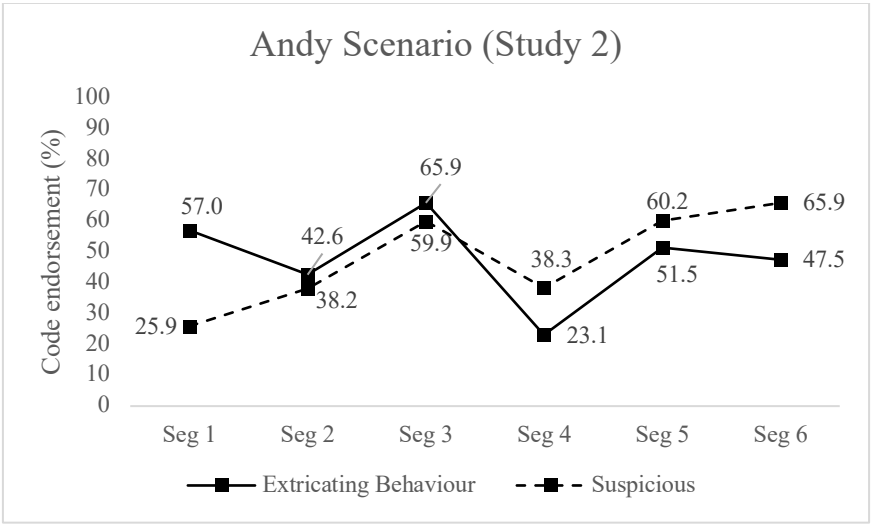
First round of inter-rater reliability for coding on each scenario (N=20)

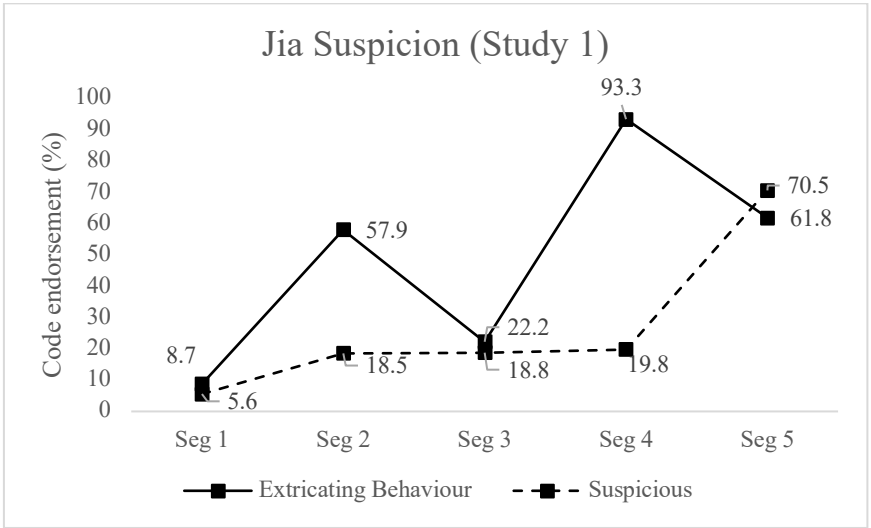
Scenario	Behaviour		Suspicion	
	% Agreement	Kappa	% Agreement	Kappa
Aaron	89	.87	82	.72
Andy	89.2	.84	76.7	.64
Amy	84	.81	81	.70
Jia	82	.75	84	.66
Casey	86	.83	82	.68
Derek	95	.93	86.3	.60

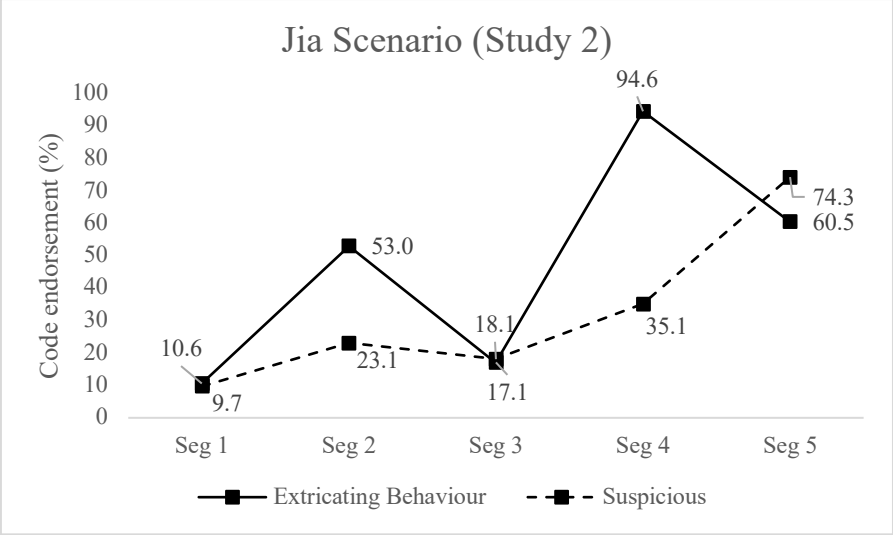
Appendix I

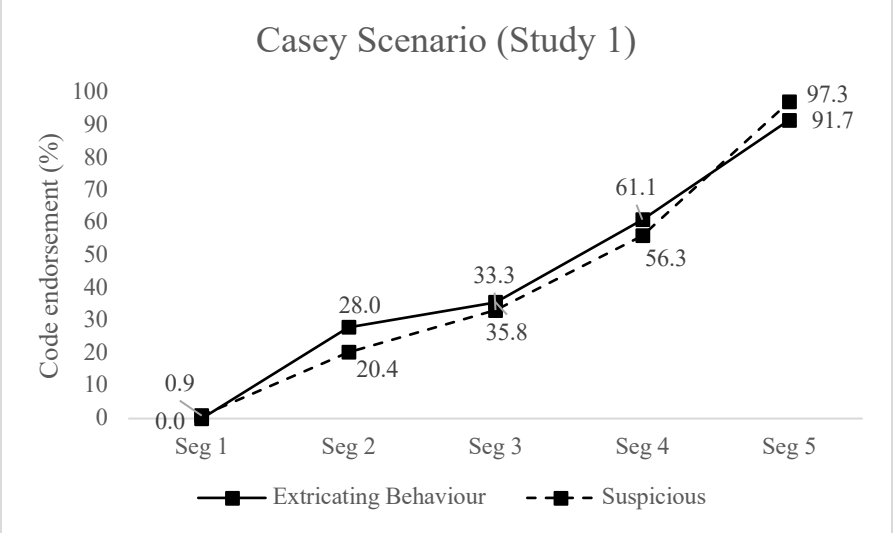
Comparison frequency graphs from Study 1 and 2

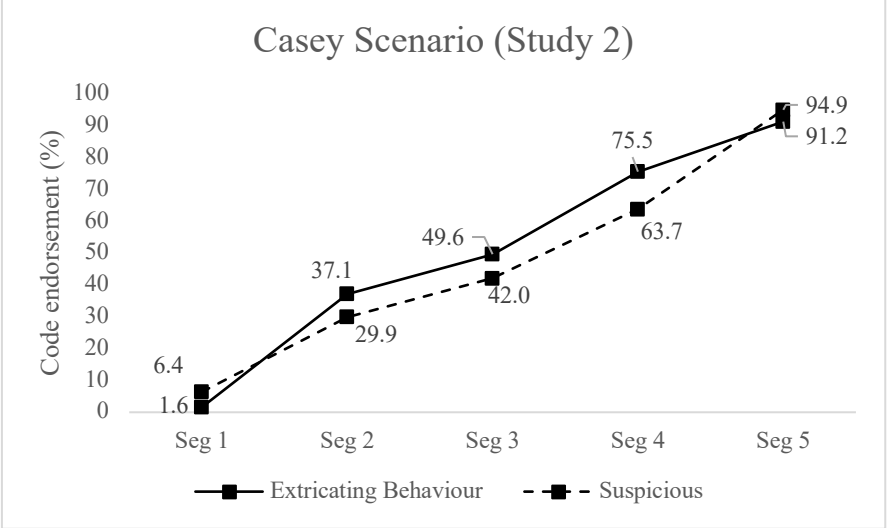


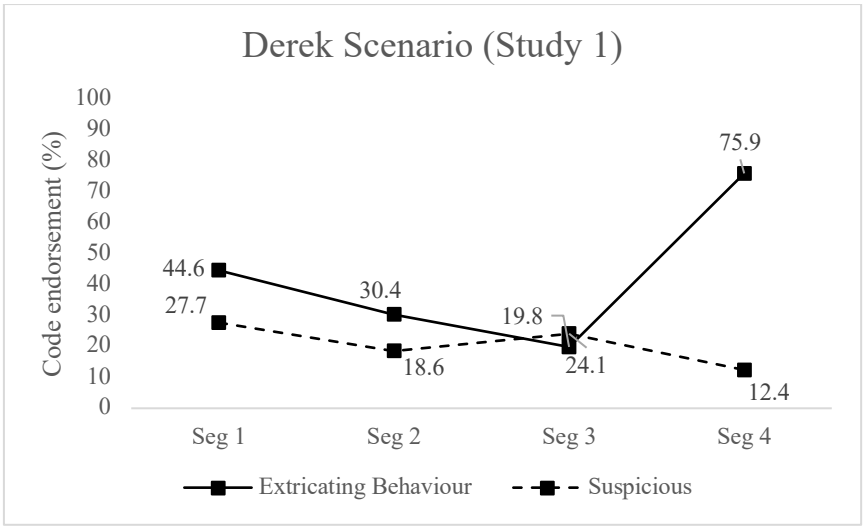


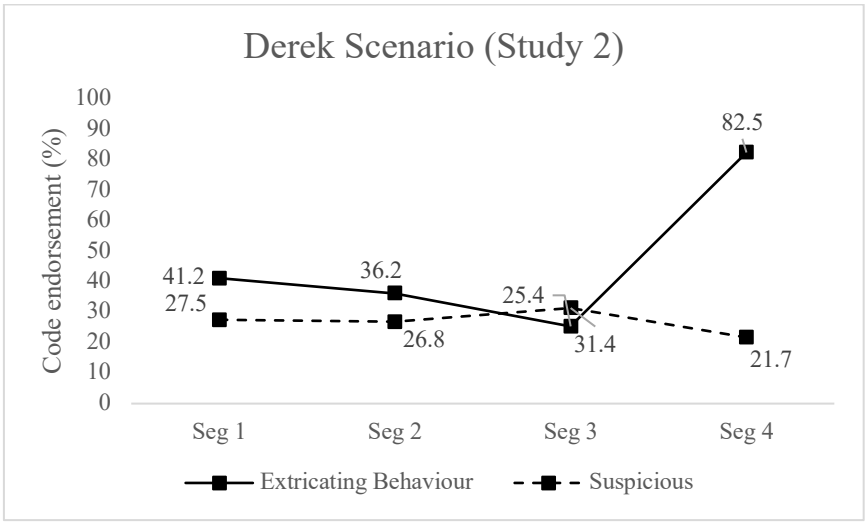



Jia Scenario (Study 2)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
10.6
9.7
53.0
23.1
18.1
17.1
94.6
35.1
74.3
60.5
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Extricating Behaviour
Suspicious


Casey Scenario (Study 1)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
0.0
0.9
28.0
20.4
33.3
35.8
61.1
56.3
97.3
91.7
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Extricating Behaviour
Suspicious


Casey Scenario (Study 2)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
1.6
6.4
37.1
29.9
49.6
42.0
75.5
63.7
94.9
91.2
Seg 1
Seg 2
Seg 3
Seg 4
Seg 5
Extricating Behaviour
Suspicious


Derek Scenario (Study 1)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
44.6
30.4
19.8
75.9
27.7
18.6
24.1
12.4
Seg 1
Seg 2
Seg 3
Seg 4
Extricating Behaviour
Suspicious


Derek Scenario (Study 2)
Code endorsement (%)
100
90
80
70
60
50
40
30
20
10
0
41.2
36.2
25.4
82.5
27.5
26.8
31.4
21.7
Seg 1
Seg 2
Seg 3
Seg 4
Extricating Behaviour
Suspicious

Appendix J

Chi-square and Cochran-Q analyses for Study 2

	N	One Sample Chi Square	*V*
Aaron Scenario			
Behaviour			
Segment 1	131	$\chi^2(1) = 4.77$, $p = .029$	.19
Segment 2	128	$\chi^2(1) = 0.50$, $p = .480$	.02
Segment 3	131	$\chi^2(1) = 11.61$, $p < .001$	.30
Segment 4	132	$\chi^2(1) = 109.09$, $p < .001$	.91
Segment 5	136	$\chi^2(1) = 47.06$, $p < .001$	.59
Suspicion			
Segment 1	138	$\chi^2(1) = 111.42$, $p < .001$	.90
Segment 2	134	$\chi^2(1) = 0.00$, $p = 1.00$	.00
Segment 3	133	$\chi^2(1) = 31.77$, $p < .001$	.49
Segment 4	137	$\chi^2(1) = 121.47$, $p < .001$	.94
Segment 5	138	$\chi^2(1) = 81.42$, $p < .001$	.77
Andy Scenario			
Behaviour			
Segment 1	135	$\chi^2(1) = 2.67$, $p = .102$	.14
Segment 2	134	$\chi^2(1) = 2.99$, $p = .084$	.15
Segment 3	132	$\chi^2(1) = 13.36$, $p < .001$	.32
Segment 4	134	$\chi^2(1) = 36.69$, $p < .001$	.52
Segment 5	132	$\chi^2(1) = 0.12$, $p = .728$	.03
Segment 6	137	$\chi^2(1) = 0.36$, $p = .550$	.05
Suspicion			
Segment 1	135	$\chi^2(1) = 31.30$, $p < .001$	.48
Segment 2	136	$\chi^2(1) = 7.53$, $p = .006$	.24
Segment 3	137	$\chi^2(1) = 5.32$, $p = .021$	.20
Segment 4	136	$\chi^2(1) = 7.53$, $p = .006$	.24
Segment 5	138	$\chi^2(1) = 5.68$, $p = .017$	.20
Segment 6	138	$\chi^2(1) = 14.03$, $p < .001$	.32
Amy Scenario			
Behaviour			
Segment 1	136	$\chi^2(1) = 34.00$, $p < .001$	.50
Segment 2	134	$\chi^2(1) = 111.08$, $p < .001$	.91
Segment 3	133	$\chi^2(1) = 42.29$, $p < .001$	.56
Segment 4	127	$\chi^2(1) = 17.39$, $p < .001$	.37
Segment 5	133	$\chi^2(1) = 129.03$, $p < .001$	.98
Suspicion			
Segment 1	137	$\chi^2(1) = 5.32$, $p = .021$	.20
Segment 2	138	$\chi^2(1) = 27.86$, $p < .001$	.45
Segment 3	138	$\chi^2(1) = 8.38$, $p = .004$	.25
Segment 4	135	$\chi^2(1) = 6.23$, $p = .013$	.21
Segment 5	137	$\chi^2(1) = 133.03$, $p < .001$	.99

	N	One Sample Chi Square	*V*
Jia Scenario			
Behaviour			
Segment 1	133	$\chi^2(1) = 82.90, p < .001$	.79
Segment 2	134	$\chi^2(1) = 0.48, p = .490$	.06
Segment 3	134	$\chi^2(1) = 57.79, p < .001$	.66
Segment 4	130	$\chi^2(1) = 103.51, p < .001$	.89
Segment 5	119	$\chi^2(1) = 5.25, p = .022$	.21
Suspicion			
Segment 1	134	$\chi^2(1) = 87.05, p < .001$	.81
Segment 2	134	$\chi^2(1) = 38.69, p < .001$	.54
Segment 3	138	$\chi^2(1) = 56.12, p < .001$	.64
Segment 4	137	$\chi^2(1) = 12.27, p < .001$	.30
Segment 5	136	$\chi^2(1) = 32.03, p < .001$	.49
Casey Scenario			
Behaviour			
Segment 1	119	$\chi^2(1) = 111.13, p < .001$	.97
Segment 2	132	$\chi^2(1) = 8.76, p = .003$	.26
Segment 3	135	$\chi^2(1) = 0.01, p = .931$	.01
Segment 4	110	$\chi^2(1) = 28.51, p < .001$	.51
Segment 5	135	$\chi^2(1) = 91.27, p < .001$	.82
Suspicion			
Segment 1	126	$\chi^2(1) = 96.03, p < .001$	.87
Segment 2	137	$\chi^2(1) = 22.08, p < .001$	.40
Segment 3	138	$\chi^2(1) = 3.51, p = .061$	.16
Segment 4	138	$\chi^2(1) = 10.46, p = .001$	.28
Segment 5	138	$\chi^2(1) = 111.42, p < .001$	.90
Derek Scenario			
Behaviour			
Segment 1	136	$\chi^2(1) = 4.24, p = .040$	.18
Segment 2	138	$\chi^2(1) = 10.46, p = .001$	.28
Segment 3	134	$\chi^2(1) = 32.51, p < .001$	.49
Segment 4	137	$\chi^2(1) = 57.82, p < .001$	.65
Suspicion			
Segment 1	138	$\chi^2(1) = 27.86, p < .001$	.45
Segment 2	138	$\chi^2(1) = 29.68, p < .001$	.46
Segment 3	137	$\chi^2(1) = 18.99, p < .001$	.37
Segment 4	138	$\chi^2(1) = 44.09, p < .001$	.57

Appendix K

Frequencies and excluded data from the SAP in Study 1 and 2

Aaron Behaviour	Valid *n*	Maladaptive	Adaptive	Missing[8]
Segment 1	236	155	81	15
Segment 2	237	117	120	14
Segment 3	236	89	147	15
Segment 4	240	14	226	11
Segment 5	242	50	192	9

Andy Behaviour	Valid *n*	Maladaptive	Adaptive	Missing
Segment 1	241	96	145	10
Segment 2	243	135	108	8
Segment 3	236	85	151	15
Segment 4	244	194	50	7
Segment 5	239	110	129	12
Segment 6	246	136	110	5

Amy Behaviour	Valid *n*	Maladaptive	Adaptive	Missing
Segment 1	249	184	65	2
Segment 2	244	8	236	7
Segment 3	241	48	193	10
Segment 4	228	72	156	23
Segment 5	243	4	239	8

Jia Behaviour	Valid *n*	Maladaptive	Adaptive	Missing
Segment 1	236	213	23	15
Segment 2	241	108	133	10
Segment 3	242	195	47	9
Segment 4	234	14	220	17
Segment 5	221	86	165	30

Casey Behaviour	Valid *n*	Maladaptive	Adaptive	Missing
Segment 1	221	219	2	30
Segment 2	239	160	79	12
Segment 3	244	138	106	7
Segment 4	200	62	138	51
Segment 5	244	21	223	7

Derek Behaviour	Valid *n*	Maladaptive	Adaptive	Missing
Segment 1	248	142	106	3
Segment 2	250	166	84	1
Segment 3	245	189	56	6
Segment 4	249	198	51	2

[8] Missing = ‘non-behavioural’, ‘other’ or missing due to audio failure.

Aaron Suspicion	Valid *n*	Not suspicious	Suspicious	Missing[9]
Segment 1	251	241	10	0
Segment 2	246	129	117	5
Segment 3	245	72	173	6
Segment 4	250	12	238	1
Segment 5	251	32	219	0

Andy Suspicion	Valid *n*	Not suspicious	Suspicious	Missing
Segment 1	246	189	57	5
Segment 2	247	162	85	4
Segment 3	249	120	129	2
Segment 4	248	164	84	3
Segment 5	251	98	153	0
Segment 6	250	96	154	1

Amy Suspicion	Valid *n*	Not suspicious	Suspicious	Missing
Segment 1	250	159	91	1
Segment 2	248	184	64	3
Segment 3	249	159	90	2
Segment 4	247	101	146	4
Segment 5	249	3	246	2

Jia Suspicion	Valid *n*	Not suspicious	Suspicious	Missing
Segment 1	242	223	19	9
Segment 2	242	191	51	9
Segment 3	250	204	46	1
Segment 4	248	178	70	3
Segment 5	248	68	180	3

Casey Suspicion	Valid *n*	Not suspicious	Suspicious	Missing
Segment 1	237	228	9	14
Segment 2	250	186	64	1
Segment 3	249	154	95	2
Segment 4	250	99	151	1
Segment 5	250	10	240	1

Derek Suspicion	Valid *n*	Not suspicious	Suspicious	Missing
Segment 1	250	181	69	1
Segment 2	251	196	58	0
Segment 3	249	179	70	2
Segment 4	251	207	44	0

[9] Missing = ‘non-behavioural’, ‘other’ or missing due to audio failure.

Appendix L

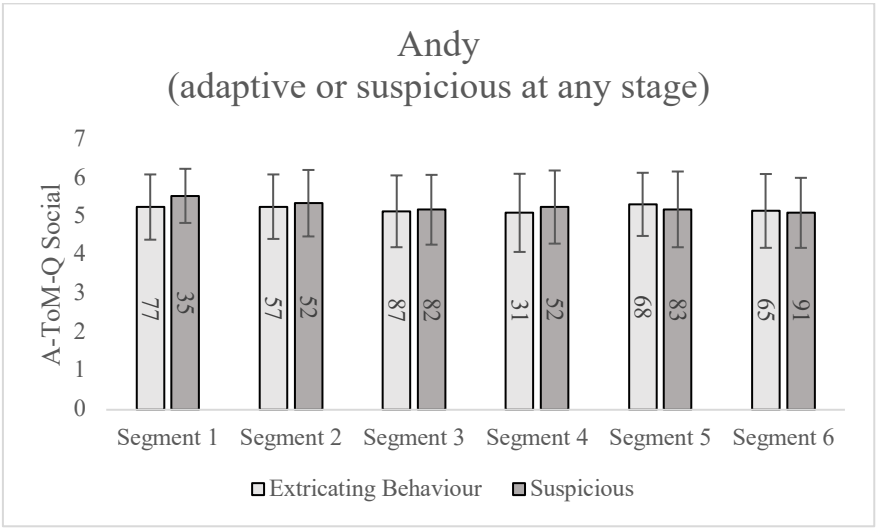
Cochran-Q and post-hoc Dunn tests with Bonferroni correction from the SAP in Study 1 and 2 combined.

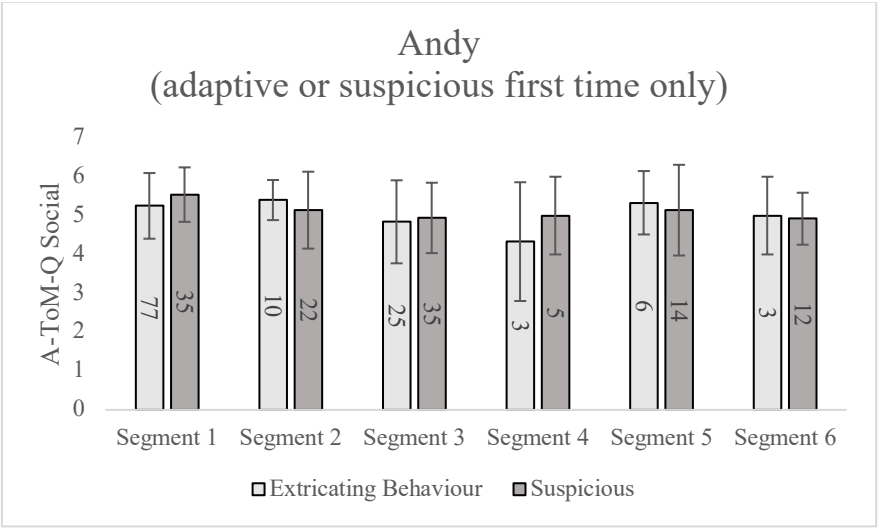
Scenario	*n*	Cochran Q test
Behaviour		
Aaron	203	χ^2 (4) = 239.35, p <.001
Andy	207	χ^2 (5) = 128.58, p <.001
Amy	213	χ^2 (4) = 353.12, p <.001
Jia	190	χ^2 (4) = 336.07, p <.001
Casey	178	χ^2 (4) = 338.68, p <.001
Derek	240	χ^2 (3) = 201.92, p <.001
Suspicion		
Aaron	241	χ^2 (4) = 565.300, p <.001
Andy	236	χ^2 (5) = 165.84, p <.001
Amy	242	χ^2 (4) = 350.54, p <.001
Jia	234	χ^2 (4) = 333.95, p <.001
Casey	235	χ^2 (4) = 470.29, p <.001
Derek	248	χ^2 (3) = 14.54, p =.002

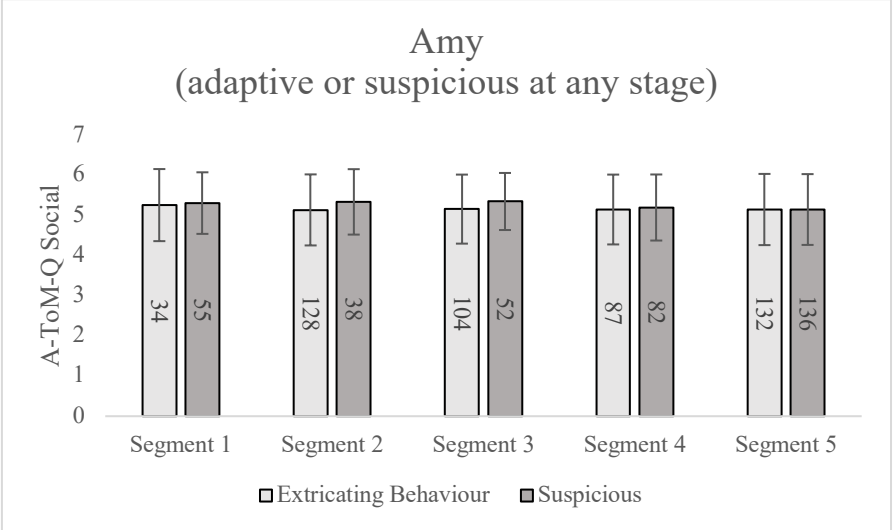
Dunn-test	Aaron	Andy	Amy	Jia	Casey	Derek
Behaviour						
Seg 1 (Seg 2)	<.001	.052	<.001	<.001	<.001	.143
Seg 2 (Seg 3)	.029	.002	.023	<.001	.125	.143
Seg 3 (Seg 4)	<.001	<.001	.005	<.001	<.001	<.001
Seg 4 (Seg 5)	.029	<.001	<.001	<.001	<.001	-
Seg 5 (Seg 6)	-	.767	-	-	-	-
Suspicion						
Seg 1 (Seg 2)	<.001	.055	.195	.008	<.001	.753
Seg 2 (Seg 3)	<.001	<.001	.317	1.00	.075	.582
Seg 3 (Seg 4)	<.001	<.001	<.001	.129	<.001	.005
Seg 4 (Seg 5)	.581	<.001	<.001	<.001	<.001	-
Seg 5 (Seg 6)	-	1.00	-	-	-	-

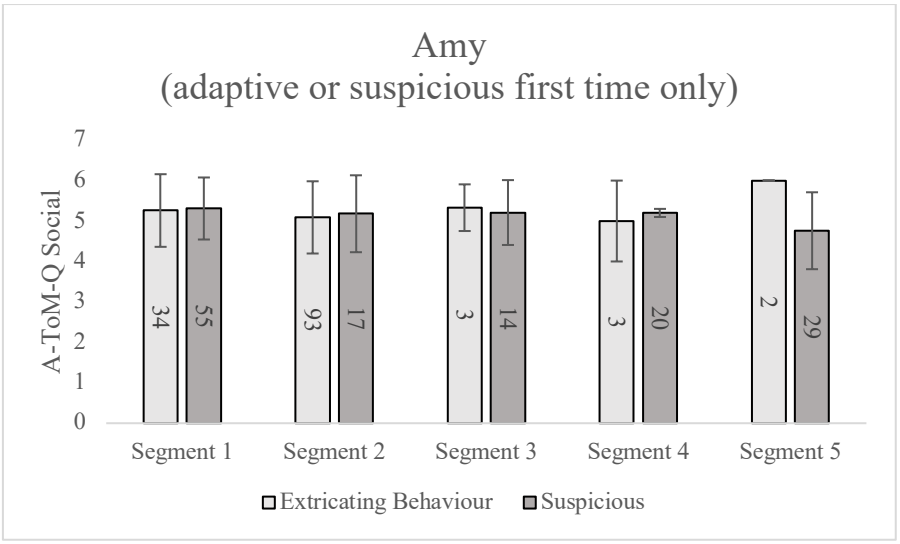
Appendix M

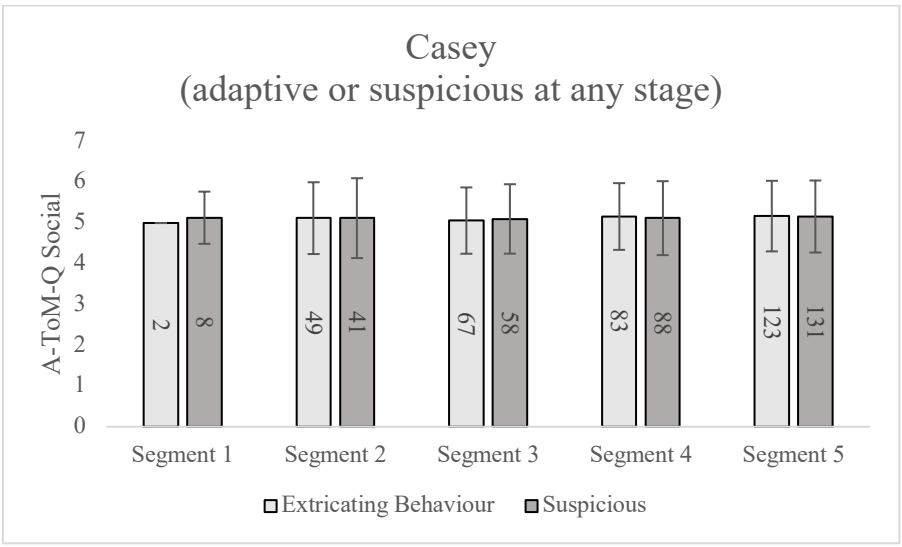
Plots showing the mean A-ToM-Q Social score (with standard deviation error bars and frequency data values) for people who responded with an adaptive behaviour or suspiciously at each segment, and for the first time only.

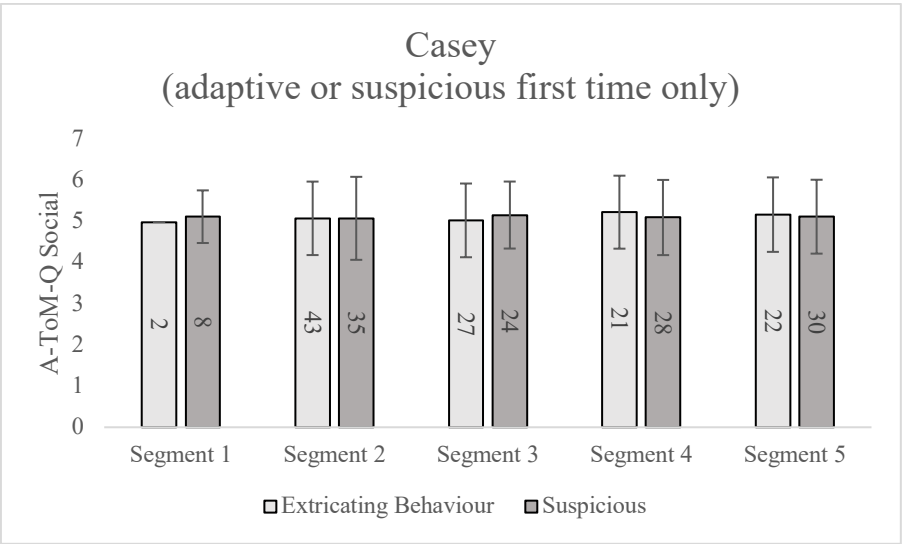


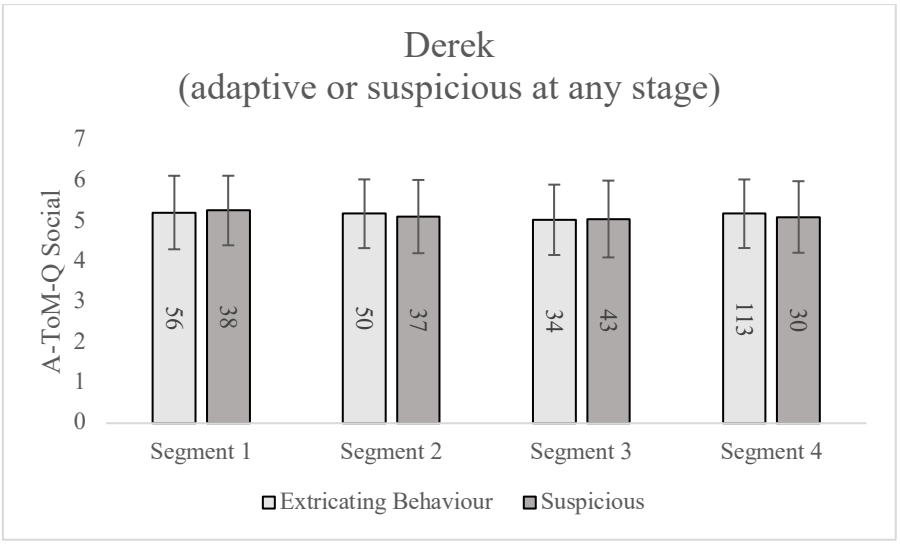


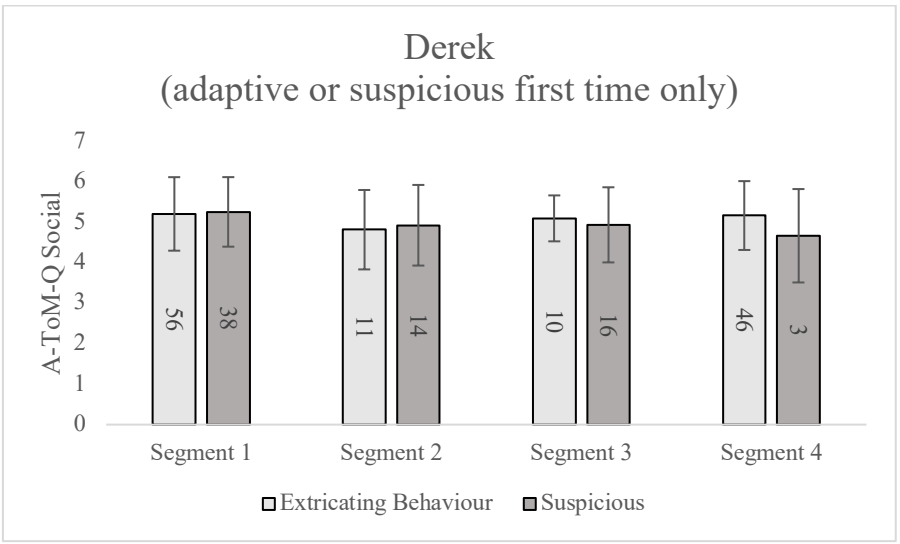



Amy
(adaptive or suspicious first time only)
A-ToM-Q Social
0
1
2
3
4
5
6
7
34
55
93
17
3
14
3
20
2
29
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicious


Casey
(adaptive or suspicious at any stage)
A-ToM-Q Social
0
1
2
3
4
5
6
7
2
8
49
41
67
58
83
88
123
131
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicious


Casey
(adaptive or suspicious first time only)
A-ToM-Q Social
0
1
2
3
4
5
6
7
2
8
43
35
27
24
21
28
22
30
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicious


Derek
(adaptive or suspicious at any stage)
A-ToM-Q Social
0
1
2
3
4
5
6
7
56
38
50
37
34
43
113
30
Segment 1
Segment 2
Segment 3
Segment 4
Extricating Behaviour
Suspicious


Derek
(adaptive or suspicious first time only)
A-ToM-Q Social
0
1
2
3
4
5
6
7
56
38
11
14
10
16
46
3
Segment 1
Segment 2
Segment 3
Segment 4
Extricating Behaviour
Suspicious

Appendix N

Point biserial correlation coefficients (Pearson) for the SAP and AQ50

	Behaviour			Suspicion		
Segment	*r*	*p*	95% CI	*r*	*p*	95% CI
Aaron						
Segment 1	.15	.086	-.02, .31	.03	.707	-.14, .20
Segment 2	.02	.817	-.15, .19	.10	.270	-.07, .26
Segment 3	-.01	.877	-.18, .16	.01	.896	-.16, .18
Segment 4	.08	.390	-.10, .24	.06	.465	-.11, .23
Segment 5	-.04	.658	-.21, .13	.01	.924	-.16, .18
Andy						
Segment 1	.01	.943	-.16, .17	.12	.152	-.05, .29
Segment 2	.00	.991	-.17, .17	.05	.553	-.12, .22
Segment 3	-.08	.355	-.25, .09	.07	.412	-.10, .24
Segment 4	.13	.123	-.04, .30	.03	.737	-.14, .20
Segment 5	-.10	.263	-.26, .07	-.04	.634	-.21, .13
Segment 6	-.09	.303	-.25, .08	-.09	.272	-.26, .07
Amy						
Segment 1	.08	.382	-.09, .24	.17	.052	.00, .32
Segment 2	-.04	.618	-.21, .13	.03	.721	-.14, .20
Segment 3	.12	.159	-.05, .29	.16	.065	-.01, .32
Segment 4	-.05	.602	-.22, .13	.04	.650	-.13, .21
Segment 5	.12	.176	-.05, .28	-.05	.590	-.21, .12
Jia						
Segment 1	.04	.622	-.13, .21	.02	.783	-.15, .19
Segment 2	-.01	.889	-.18, .16	.05	.550	-.12, .22
Segment 3	-.02	.797	-.19, .15	.03	.685	-.13, .20
Segment 4	-.05	.575	-.22, .12	.03	.705	-.14, .20
Segment 5	-.01	.917	-.19, .17	.00	.989	-.17, .17
Casey						
Segment 1	.17	.069	-.01, .34	.04	.658	-.14, .21
Segment 2	.05	.550	-.12, .22	.09	.295	-.08, .25
Segment 3	.14	.105	-.03, .30	.24	.005	.07, .39
Segment 4	.24	.011	.06, .41	.26	.002	.10, .41
Segment 5	-.12	.171	-.28, .05	-.05	.594	-.21, .12
Derek						
Segment 1	-.15	.089	-.31, .02	-.10	.242	-.26, .07
Segment 2	-.01	.868	-.18, .15	-.05	.546	-.22, .12
Segment 3	-.02	.816	-.19, .15	.04	.676	-.13, .20
Segment 4	.02	.813	-.15, .19	.05	.564	-.12, .21

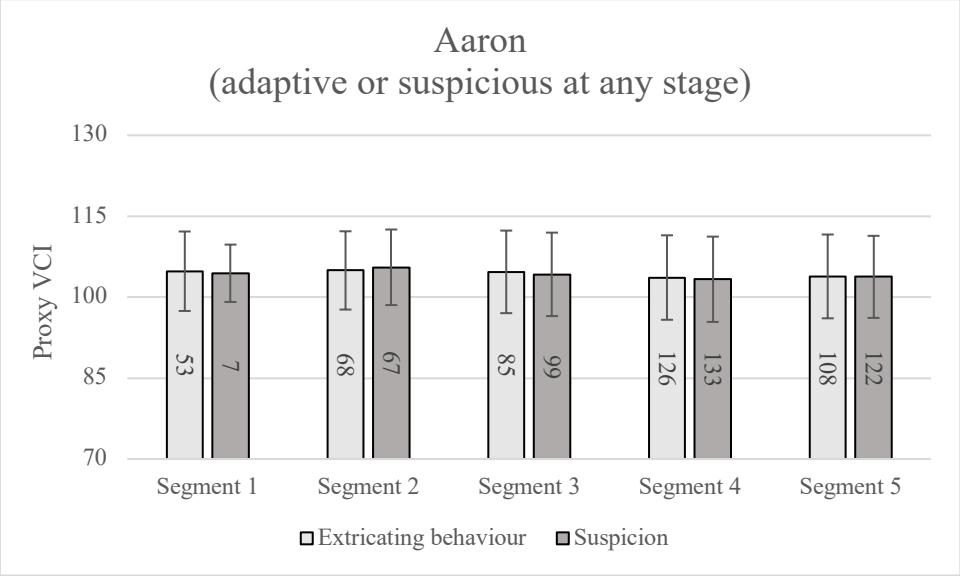
Appendix O

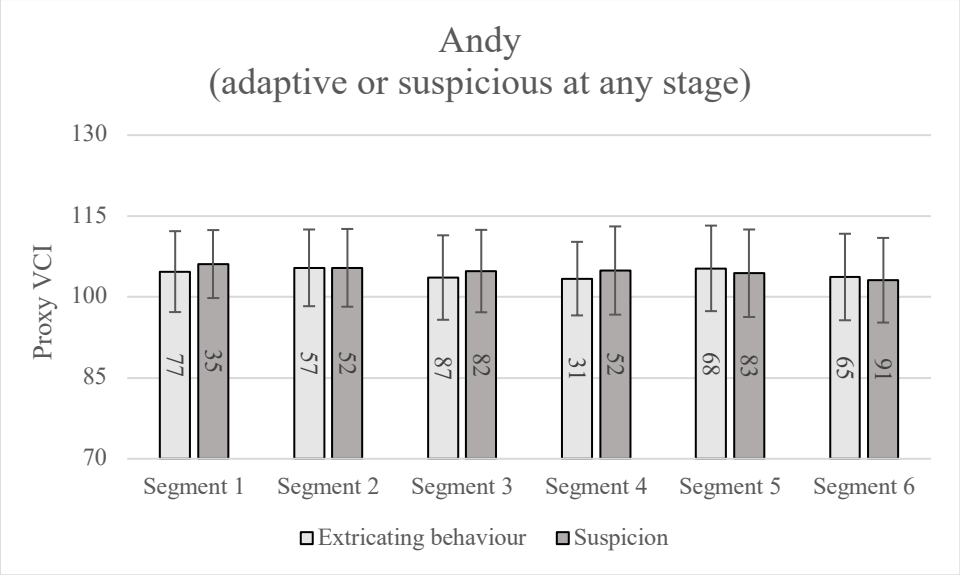
Point biserial correlation coefficients (Pearson) for the SAP and VCI for participants who scored 85 or over (6 exclusions)

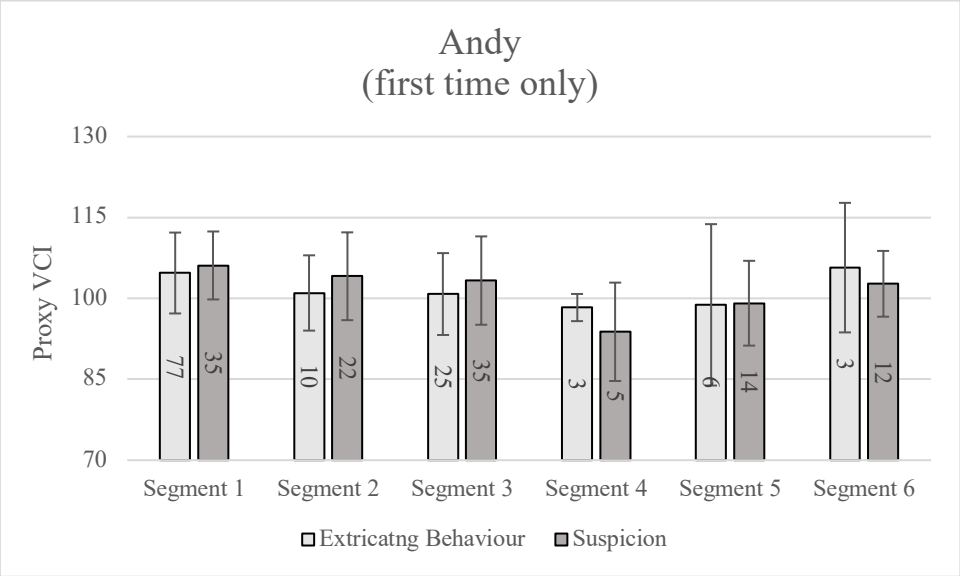
	Behaviour			Suspicion		
Segment	*r*	*p*	95% CI	*r*	*p*	95% CI
Aaron						
Segment 1	.11	.230	-.07, .28	.01	.918	-.16, .18
Segment 2	.18	.047	.00, .35	.26	.003	.09, .41
Segment 3	.25	.005	.08, .41	.20	.023	.03, .36
Segment 4	.15	.091	-.02, .32	.00	.971	-.17, .17
Segment 5	.15	.092	-.02, .31	.08	.347	-.09, .25
Andy						
Segment 1	.13	.150	-.05, .29	.17	.060	-.01, .33
Segment 2	.23	.008	.06, .39	.16	.061	-.01, .33
Segment 3	.03	.728	-.14, .20	.16	.070	-.01, .32
Segment 4	-.01	.892	-.18, .16	.24	.006	.07, .40
Segment 5	.28	.001	.12, .44	.21	.017	.04, .37
Segment 6	.01	.906	-.16, .18	-.07	.458	-.23, .11
Amy						
Segment 1	-.01	.936	-.18, .16	.00	.992	-.17, .17
Segment 2	-.07	.413	-.24, .10	.04	.685	-.14, .21
Segment 3	.15	.082	-.02, .32	-.01	.937	-.18, .16
Segment 4	.11	.245	-.07, .28	.12	.172	-.05, .29
Segment 5	-.02	.799	-.20, .15	-	-	-
Jia						
Segment 1	.12	.188	-.06, .29	.21	.017	.04, .37
Segment 2	.10	.244	-.07, .27	.18	.037	.01, .34
Segment 3	.07	.446	-.11, .24	.21	.015	.04, .37
Segment 4	.12	.187	-.06, .29	.05	.563	-.12, .22
Segment 5	.07	.488	-.12, .25	.12	.167	-.05, .29
Casey						
Segment 1	.13	.152	-.05, .31	.09	.334	-.09, .26
Segment 2	-.10	.256	-.27, .07	-.17	.056	-.33, .00
Segment 3	-.01	.919	-.18, .16	.01	.865	-.16, .19
Segment 4	.00	.977	-.19, .19	.07	.431	-.10, .24
Segment 5	.05	.550	-.12, .22	.06	.467	-.11, .23
Derek						
Segment 1	.26	.002	.10, .42	.18	.036	.01, .34
Segment 2	.23	.009	.06, .38	.18	.037	.01, .34
Segment 3	.10	.275	-.08, .27	.12	.182	-.06, .28
Segment 4	-.08	.356	-.25, .09	.14	.112	-.03, .30

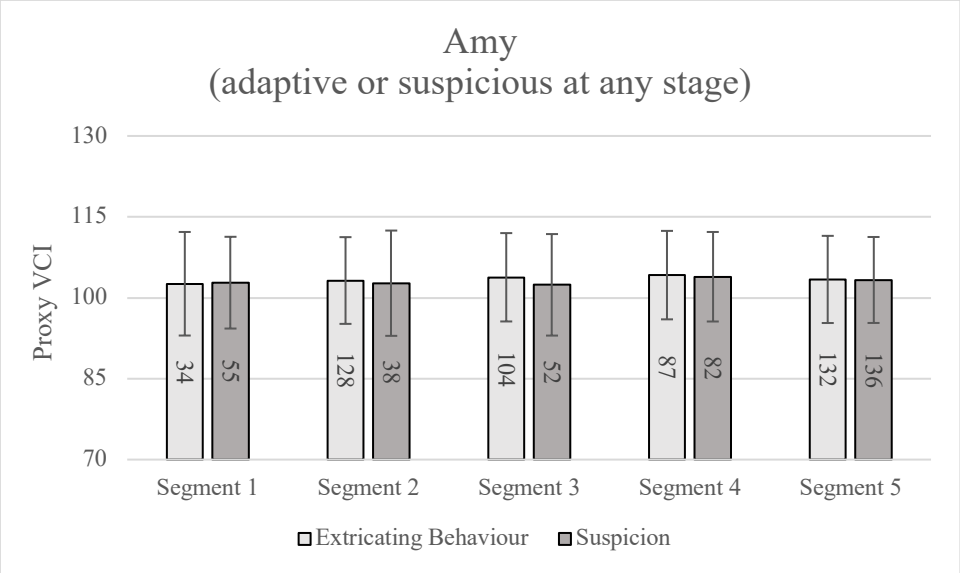
Appendix P

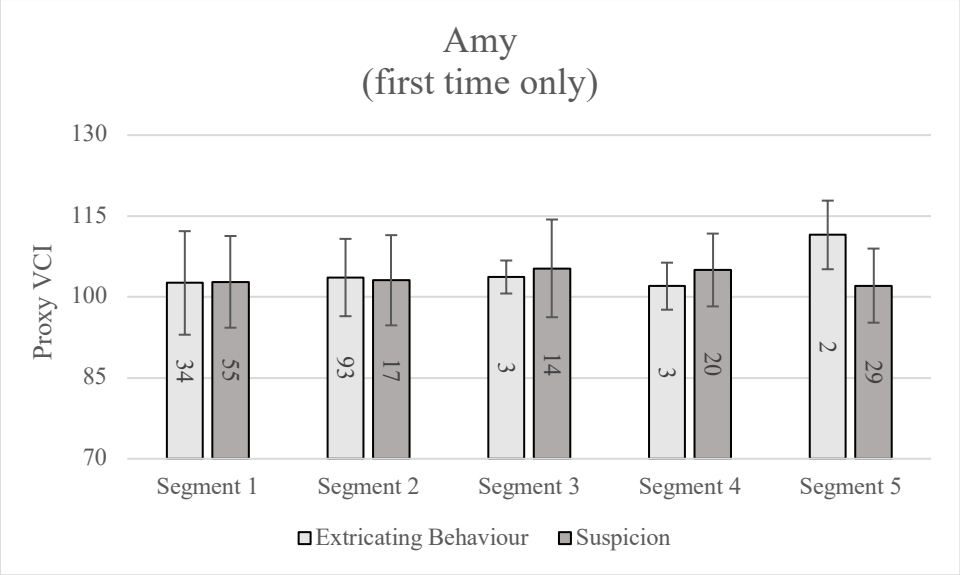
Plots showing the mean proxy VCI score (with standard deviation error bars and frequency data values) for people who responded with an adaptive behaviour or suspiciously at each segment, and for the first time only.

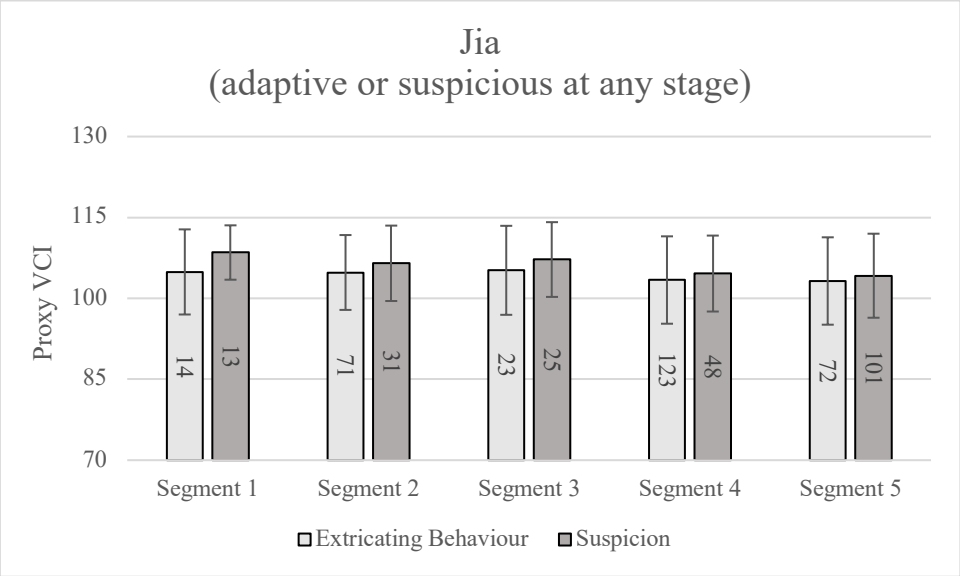


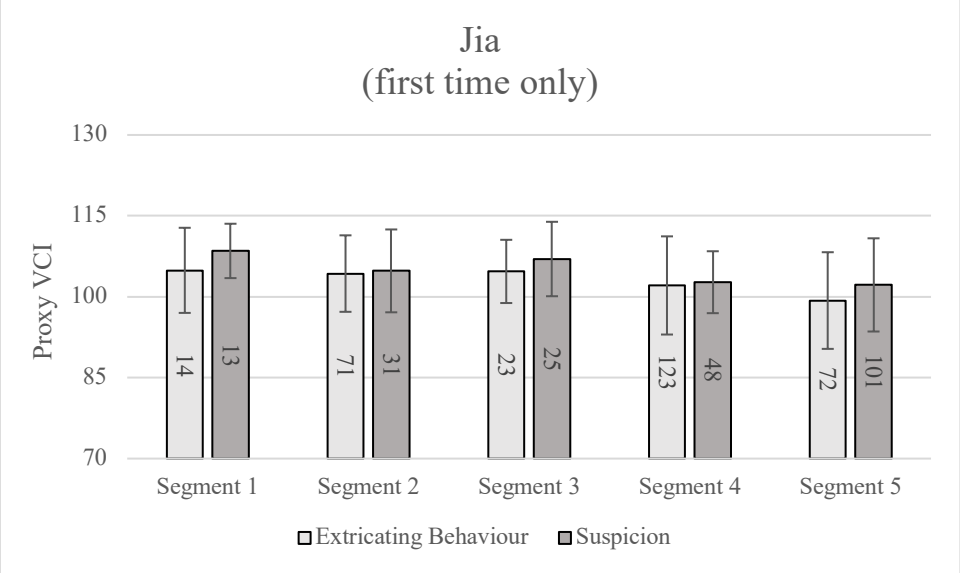

Andy
(adaptive or suspicious at any stage)
130
115
100
85
70
Proxy VCI
77
35
57
52
87
82
31
52
68
83
65
91
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Segment 6
Extricating behaviour
Suspicion

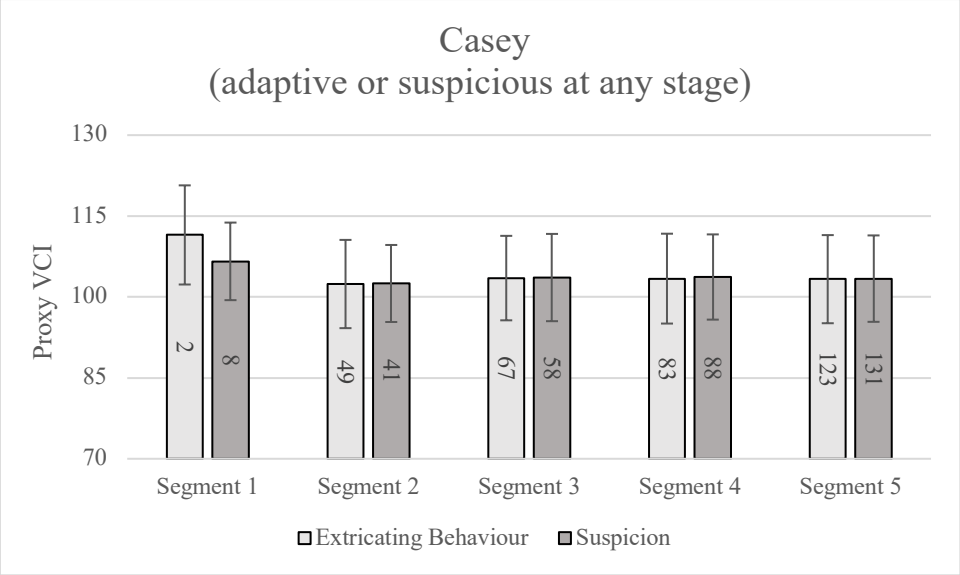

Andy
(first time only)
130
115
100
85
70
Proxy VCI
77
35
10
22
25
35
3
5
6
14
3
12
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Segment 6
Extricatng Behaviour
Suspicion

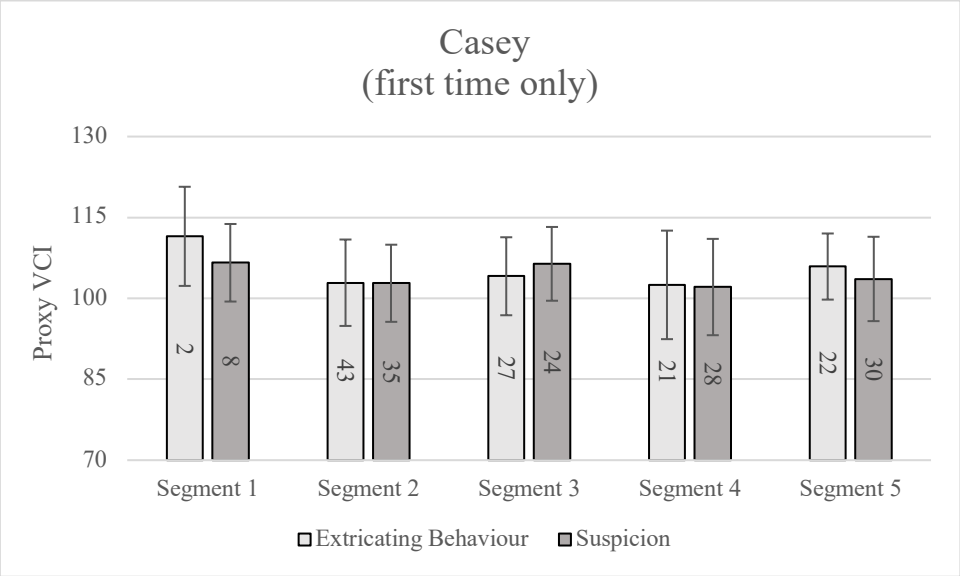

Amy
(adaptive or suspicious at any stage)
130
115
100
85
70
Proxy VCI
34
55
128
38
104
52
87
82
132
136
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicion

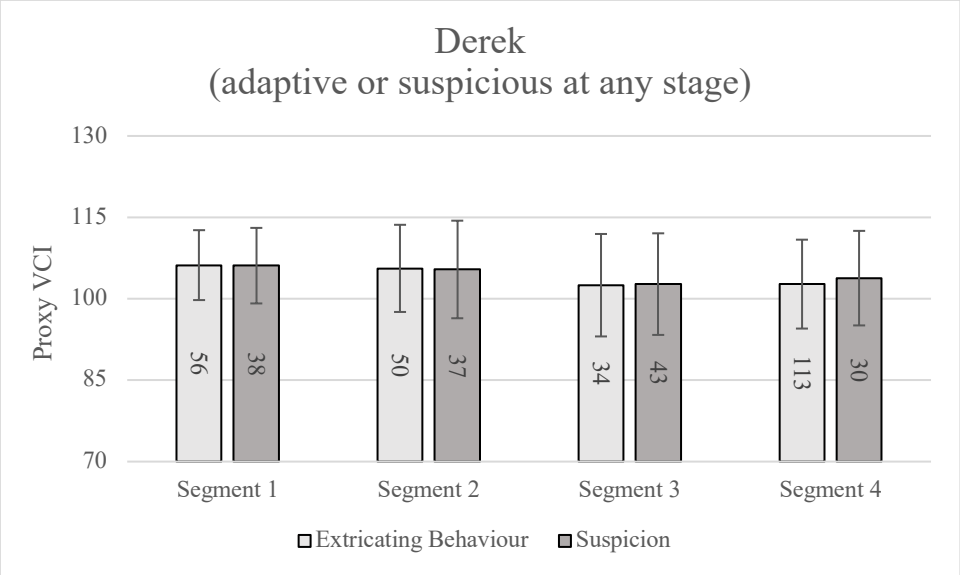

Amy
(first time only)
130
115
100
85
70
Proxy VCI
34
55
93
17
3
14
3
20
2
29
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicion

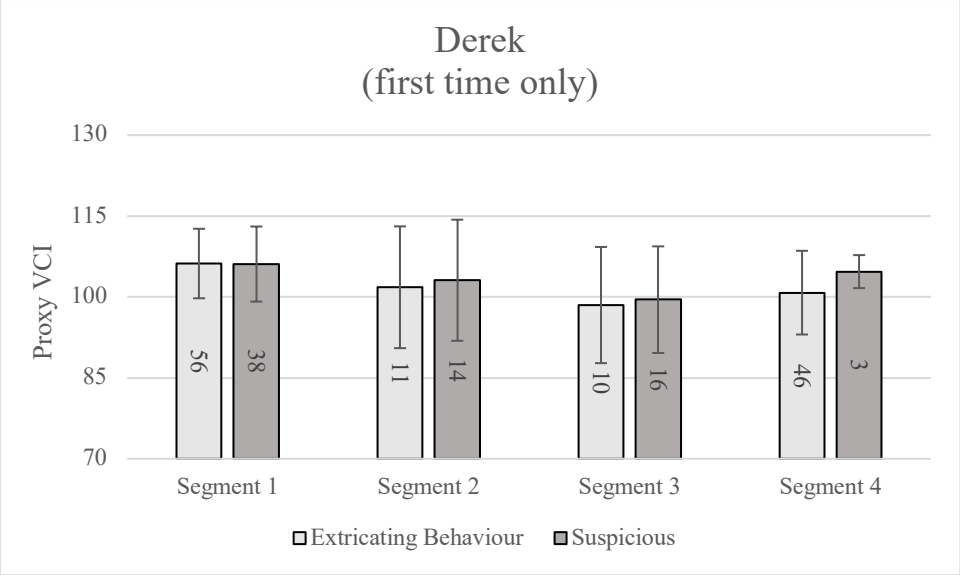

Jia
(adaptive or suspicious at any stage)
130
115
100
85
70
Proxy VCI
14
13
71
31
23
25
123
48
72
101
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicion


Jia
(first time only)
130
115
100
85
70
Proxy VCI
14
13
71
31
23
25
123
48
72
101
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicion


Casey
(adaptive or suspicious at any stage)
Proxy VCI
130
115
100
85
70
2
8
49
41
67
58
83
88
123
131
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicion


Casey
(first time only)
Proxy VCI
130
115
100
85
70
2
8
43
35
27
24
21
28
22
30
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Extricating Behaviour
Suspicion


Derek
(adaptive or suspicious at any stage)
Proxy VCI
130
115
100
85
70
56
38
50
37
34
43
113
30
Segment 1
Segment 2
Segment 3
Segment 4
Extricating Behaviour
Suspicion


Derek
(first time only)
Proxy VCI
130
115
100
85
70
56
38
11
14
10
16
46
3
Segment 1
Segment 2
Segment 3
Segment 4
Extricating Behaviour
Suspicious

Appendix Q

Chi-square analyses comparing SAP responses from the 'pseudo' normative sample (Study 1 and 2) and Study 3.

Aaron Scenario	N	Chi Square	φ
Behaviour			
Segment 1	428	$\chi^2(1) = 3.97$, $p = .046$	.10
Segment 2	429	$\chi^2(1) = 7.79$, $p = .005$	.14
Segment 3	430	$\chi^2(1) = 2.51$, $p = .113$	.08
Segment 4	436	$\chi^2(1) = 0.11$, $p = .739$	.02
Segment 5	441	$\chi^2(1) = 1.19$, $p = .276$	.05
Suspicion			
Segment 1	452	$\chi^2(1) = 3.26$, $p = .071$	.09
Segment 2	442	$\chi^2(1) = 11.58$, $p < .001$	.16
Segment 3	445	$\chi^2(1) = 4.59$, $p = .032$	.10
Segment 4	451	$\chi^2(1) = 0.30$, $p = .582$	-.03
Segment 5	454	$\chi^2(1) = 2.23$, $p = .135$	.07

Andy Scenario	N	One Sample Chi Square	φ
Behaviour			
Segment 1	434	$\chi^2(1) = 0.04$, $p = .837$	.01
Segment 2	442	$\chi^2(1) = 5.14$, $p = .023$	.11
Segment 3	423	$\chi^2(1) = 0.04$, $p = .852$	-.01
Segment 4	443	$\chi^2(1) = 1.97$, $p = .161$	.07
Segment 5	435	$\chi^2(1) = 0.20$, $p = .654$	.02
Segment 6	445	$\chi^2(1) = 0.41$, $p = .525$	-.03
Suspicion			
Segment 1	442	$\chi^2(1) = 0.66$, $p = .416$	.04
Segment 2	448	$\chi^2(1) = 4.10$, $p = .043$	.10
Segment 3	447	$\chi^2(1) = 5.77$, $p = .016$	.11
Segment 4	450	$\chi^2(1) = 1.86$, $p = .173$	.06
Segment 5	454	$\chi^2(1) = 0.04$, $p = .853$	-.01
Segment 6	452	$\chi^2(1) = 0.24$, $p = .621$	.02

Amy Scenario	N	One Sample Chi Square	*V*
Behaviour			
Segment 1	446	$\chi^2(1) = 2.53, p = .112$	.08
Segment 2	442	$\chi^2(1) = 3.32, p = .068$	-.09
Segment 3	438	$\chi^2(1) = 0.38, p = .537$	-.03
Segment 4	418	$\chi^2(1) = 0.67, p = .413$	.04
Segment 5	443	$\chi^2(1) = 4.98, p = .026$	-.11
Suspicion			
Segment 1	451	$\chi^2(1) = 6.99, p = .008$	.12
Segment 2	451	$\chi^2(1) = 0.10, p = .758$	.02
Segment 3	451	$\chi^2(1) = 1.48, p = .223$	-.06
Segment 4	447	$\chi^2(1) = 0.53, p = .466$	.04
Segment 5	451	$\chi^2(1) = 3.56, p = .059$	-.09

Jia Scenario	N	One Sample Chi Square	φ
Behaviour			
Segment 1	431	$\chi^2(1) = 14.28, p < .001$	.18
Segment 2	439	$\chi^2(1) = 0.06, p = .810$	-.01
Segment 3	439	$\chi^2(1) = 22.70, p < .001$	.23
Segment 4	427	$\chi^2(1) = 0.15, p = .901$	.01
Segment 5	393	$\chi^2(1) = 0.23, p = .635$	-.02
Suspicion			
Segment 1	440	$\chi^2(1) = 8.93, p = .003$	.14
Segment 2	443	$\chi^2(1) = 4.50, p = .034$	.10
Segment 3	452	$\chi^2(1) = 11.45, p < .001$	.16
Segment 4	447	$\chi^2(1) = 1.03, p = .310$	.05
Segment 5	444	$\chi^2(1) = 0.33, p = .566$	.027

Derek Scenario	N	One Sample Chi Square	φ
Behaviour			
Segment 1	449	$\chi^2(1) = 4.90, p = .027$	.10
Segment 2	450	$\chi^2(1) = 3.75, p = .053$	.09
Segment 3	447	$\chi^2(1) = 3.93, p = .047$	.09
Segment 4	448	$\chi^2(1) = 3.13, p = .077$	.08
Suspicion			
Segment 1	453	$\chi^2(1) = 1.56, p = .212$	.06
Segment 2	452	$\chi^2(1) = 9.46, p = .002$	.15
Segment 3	451	$\chi^2(1) = 1.10, p = .294$	.05
Segment 4	454	$\chi^2(1) = 1.86, p = .173$	.06

Appendix R

		ASD group	N	Mean	SD	95% CI
Aaron	Segment 1	non-ASD	94	8.76	1.66	8.41, 9.10
		ASD	102	8.43	1.77	8.08, 8.78
	Segment 2	non-ASD	94	8.96	1.66	8.62, 9.30
		ASD	102	8.74	1.65	8.41, 9.06
	Segment 3	non-ASD	94	8.97	1.56	8.65, 9.29
		ASD	102	8.95	1.78	8.60, 9.30
	Segment 4	non-ASD	94	8.96	1.77	8.60, 9.32
		ASD	102	9.28	1.37	9.02, 9.55
	Segment 5	non-ASD	94	9.07	1.52	8.76, 9.39
		ASD	107	9.16	1.41	8.88, 9.43
Andy	Segment 1	non-ASD	94	8.74	1.82	8.37, 9.12
		ASD	102	8.50	1.99	8.11, 8.89
	Segment 2	non-ASD	94	8.67	1.48	8.37, 8.97
		ASD	102	8.36	1.82	8.00, 8.72
	Segment 3	non-ASD	94	8.70	1.52	8.39, 9.01
		ASD	102	8.18	2.22	7.74, 8.61
	Segment 4	non-ASD	94	8.57	1.78	8.21, 8.94
		ASD	102	8.27	2.13	7.86, 8.69
	Segment 5	non-ASD	94	8.72	1.52	8.41, 9.03
		ASD	102	8.48	1.98	8.09, 8.87
	Segment 6	non-ASD	94	8.31	1.52	8.00, 8.62
		ASD	102	8.31	1.90	7.94, 8.69
Amy	Segment 1	non-ASD	94	8.94	1.42	8.65, 9.23
		ASD	102	8.47	1.94	8.09, 8.85
	Segment 2	non-ASD	94	8.84	1.54	8.52, 9.16
		ASD	102	8.71	1.81	8.35, 9.06
	Segment 3	non-ASD	94	8.50	1.92	8.11, 8.89
		ASD	102	8.69	1.76	8.34, 9.03
	Segment 4	non-ASD	94	8.77	1.69	8.42, 9.11
		ASD	102	8.78	1.83	8.43, 9.14

	Segment 5	non-ASD	94	9.21	1.31	8.94, 9.48
		ASD	102	9.04	1.47	8.75, 9.33
Jia	Segment 1	non-ASD	94	8.79	1.56	8.47, 9.11
		ASD	102	8.82	1.50	8.53, 9.12
	Segment 2	non-ASD	94	8.63	1.51	8.32, 8.94
		ASD	102	8.54	1.57	8.23, 8.85
	Segment 3	non-ASD	94	8.61	1.77	8.24, 8.97
		ASD	102	8.66	1.61	8.34, 8.97
	Segment 4	non-ASD	94	8.89	1.49	8.59, 9.20
		ASD	102	8.70	1.82	8.34, 9.05
	Segment 5	non-ASD	94	8.86	1.74	8.51, 9.22
		ASD	102	8.51	2.25	8.07, 8.95
Derek	Segment 1	non-ASD	94	8.90	1.41	8.62, 9.19
		ASD	102	8.91	1.50	8.62, 9.21
	Segment 2	non-ASD	94	8.94	1.33	8.66, 9.21
		ASD	102	8.61	1.93	8.23, 8.99
	Segment 3	non-ASD	94	9.00	1.36	8.72, 9.28
		ASD	102	8.88	1.48	8.59, 9.17
	Segment 4	non-ASD	94	9.30	1.16	9.06, 9.54
		ASD	102	9.17	1.24	8.92, 9.41

	Independent Samples *t* test	Cohen's *d*	95% CI on *d*
Aaron Segment 1	*t*(194)= 1.32, *p*=.189	.19	-.09, .47
Aaron Segment 2	*t*(194)= 0.94, *p*=.349	.13	-.15, .41
Aaron Segment 3	*t*(194)= 0.07, *p*=.472	.01	-.27, .29
Aaron Segment 4	*t*(174.99)= -1.44, *p*=.151	-.21	-.49, .07
Aaron Segment 5	*t*(194)= -0.40, *p*=.694	-.06	-.34, .22
Andy Segment 1	*t*(194)= 0.90, *p*=.371	.13	-.15, .41
Andy Segment 2	*t*(194)= 1.29, *p*=.198	.18	-.10, .47
Andy Segment 3	*t*(179.69)= 1.95, *p*=.053	.27	-.01, .56
Andy Segment 4	*t*(194)= 1.07, *p*=.288	.15	-.13, .43
Andy Segment 5	*t(*194)= 0.96, *p*=.339	.14	-.14, .42
Andy Segment 6	*t(*194)= -0.02, *p*=.983	.00	-.28, .28
Amy Segment 1	*t(*194)= 0.96, *p*=.339	.27	-.01, .55
Amy Segment 2	*t(*194)= 0.56, *p*=.289	.08	-.20, .36
Amy Segment 3	*t(*194)= -0.71, *p*=.479	-.10	-.38, .18
Amy Segment 4	*t(*194)= -0.07, *p*=.942	-.01	-.29, .27
Amy Segment 5	*t(*194)= 0.87, *p*=.385	.12	-.16, .40
Jia Segment 1	*t(*194)= -0.17, *p*=.868	-.02	-.30, .26
Jia Segment 2	*t(*194)= 0.40, *p*=.689	.06	-.22, .34
Jia Segment 3	*t(*194)= -0.21, *p*=.835	-.03	-.31, .25
Jia Segment 4	*t(*194)= 0.83, *p*=.409	.12	-.16, .40
Jia Segment 5	*t(*188.51)= 1.23, *p*=.220	.17	-.11, .45
Derek Segment 1	*t(*194)= -0.04, *p*=.971	-.01	-.29, .28
Derek Segment 2	*t(*180.27)= 1.39, *p*=.165	.20	-.08, .48
Derek Segment 3	*t(*194)= 0.56, *p*=.565	.08	-.20, .36
Derek Segment 4	*t(*202)= 0.76, *p*=.447	.11	-.17, .39

Mann-Whitney U	Median		*U*	*p*
	Non-ASD	ASD		
Aaron Segment 1	9	9	4153.00	.091
Aaron Segment 2	10	10	4404.00	.283
Aaron Segment 3	10	10	4977.50	.606
Aaron Segment 4	10	10	5196.50	.227
Aaron Segment 5	10	10	4981.50	.584
Andy Segment 1	9	9	4530.00	.478
Andy Segment 2	9	9	4464.50	.387
Andy Segment 3	9	9	4362.50	.256
Andy Segment 4	9	9	4471.00	.393
Andy Segment 5	9	9	4629.50	.663
Andy Segment 6	8	9	5047.50	.511
Amy Segment 1	10	9	4248.50	.143
Amy Segment 2	9	9	4678.00	.755
Amy Segment 3	9	10	5019.50	.543
Amy Segment 4	10	10	4852.50	.873
Amy Segment 5	10	10	4466.00	.348
Jia Segment 1	9	9	4789.50	.992
Jia Segment 2	9	9	4642.50	.689
Jia Segment 3	9	9	4755.50	.918
Jia Segment 4	9.5	9.5	4642.50	.681
Jia Segment 5	10	10	4538.50	.485
Derek Segment 1	9	9.5	4930.50	.713
Derek Segment 2	9	9	4529.50	.478
Derek Segment 3	10	10	4662.00	.717
Derek Segment 4	10	10	4483.50	.375

Appendix S

Point biserial correlation coefficients (Spearman) between SAP confidence ratings and adaptive behaviour or suspicion for the non-autistic sample

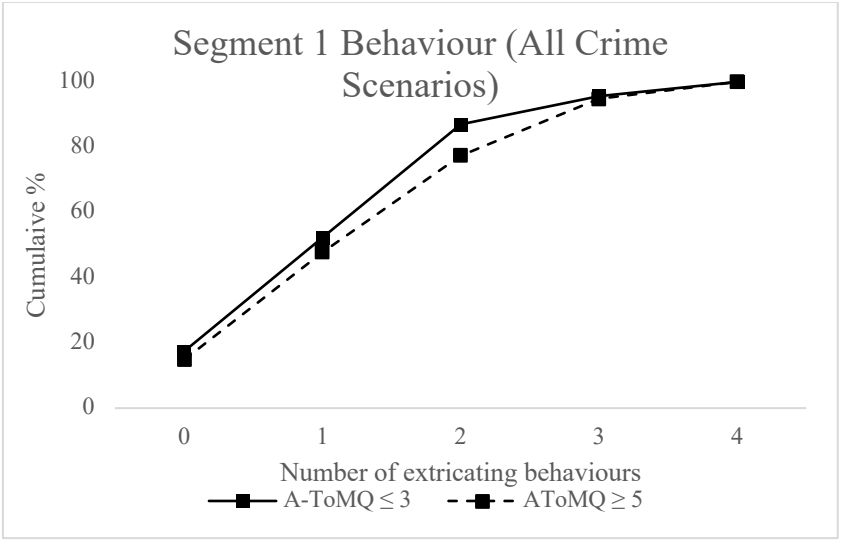
	Behaviour			Suspicion		
Segment	*Rho*	*p*	95% CI	*Rho*	*p*	95% CI
Aaron						
Segment 1	.04	.734	-.18, .25	-.07	.495	-.28, .14
Segment 2	.18	.090	-.03, .37	.21	.041	.00, .41
Segment 3	.23	.031	.02, .42	-.03	.763	-.24, .18
Segment 4	.31	.003	.10, .49	.35	<.001	.15, .52
Segment 5	.34	.001	.14, .51	.15	.152	-.06, .35
Andy						
Segment 1	.16	.134	-.06, .36	-.03	.762	-.24, .18
Segment 2	.17	.104	-.04, .37	.12	.272	-.10, .32
Segment 3	.12	.281	-.10, .32	-.01	.922	-.22, .20
Segment 4	.00	.978	-.21, .21	-.07	.490	-.28, .14
Segment 5	.04	.727	-.18, .25	.01	.907	-.20, .22
Segment 6	-.05	.644	-.26, .16	.06	.555	-.15, .267
Amy						
Segment 1	-.07	.488	-.28, .14	.14	.188	-.07, .34
Segment 2	.05	.631	-.16, .26	.15	.152	-.06, .35
Segment 3	.35	<.001	.15, .53	-.13	.230	-.33, .09
Segment 4	-.05	.635	-.26, .16	.04	.696	-.17, .25
Segment 5	.01	.914	-.20, .22	.26	.010	.06, .45
Jia						
Segment 1	-.07	.494	-.28, .14	-.06	.594	-.26, .16
Segment 2	.05	.648	-.17, .26	.02	.837	-.19, .23
Segment 3	.16	.147	-.06, .36	.00	.969	-.21, .21
Segment 4	.17	.113	-.05, .37	.10	.352	-.11, .30
Segment 5	-.11	.347	-.33, .12	-.02	.887	-.23, .20
Derek						
Segment 1	.06	.541	-.15, .27	-.08	.474	-.28, .14
Segment 2	.04	.731	-.18, .24	-.04	.737	-.24, .18
Segment 3	-.18	.084	-.37, .03	-.17	.100	-.37, .04
Segment 4	.08	.476	-.14, .28	-.13	.222	-.33, .08

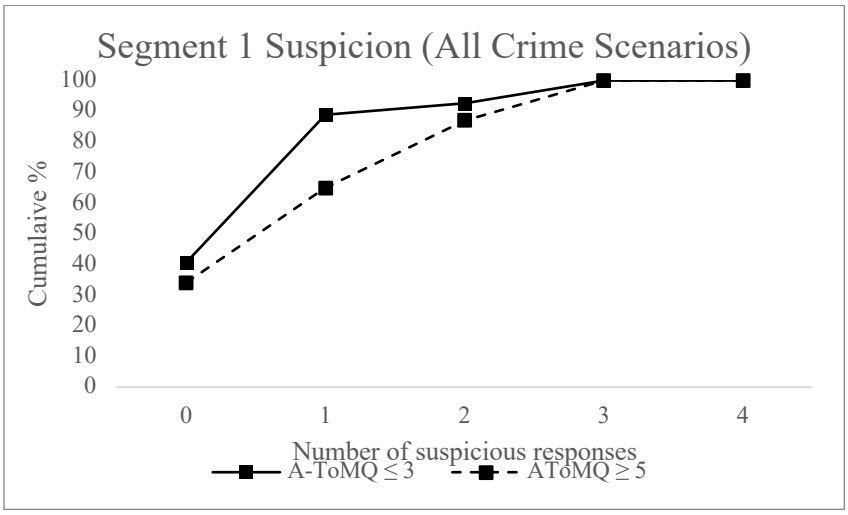
Point biserial correlation coefficients (Spearman) between SAP confidence ratings and adaptive behaviour or suspicion for the autistic sample

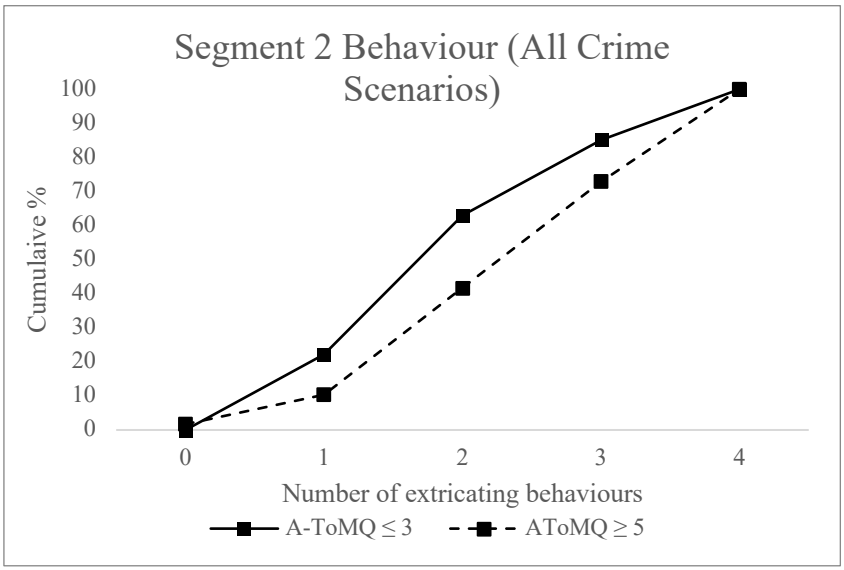
	Behaviour			Suspicion		
Segment	*Rho*	*p*	95% CI	*Rho*	*p*	95% CI
Aaron						
Segment 1	.10	.340	-.11, .30	.12	.221	-.08, .32
Segment 2	.07	.511	-.14, .27	.09	.379	-.12, .29
Segment 3	.12	.247	-.09, .31	.04	.704	-.16, .24
Segment 4	.15	.138	-.05, .34	.06	.576	-.15, .25
Segment 5	.15	.133	-.05, .34	-.01	.920	-.21, .19
Andy						
Segment 1	.09	.386	-.12, .29	-.04	.696	-.24, .17
Segment 2	.14	.179	-.07, .33	-.07	.485	-.27, .13
Segment 3	.02	.869	-.20, .23	-.19	.068	-.38, .02
Segment 4	.12	.218	-.08, .32	-.17	.087	-.36, .03
Segment 5	.03	.753	-.17, .23	-.08	.405	-.28, .12
Segment 6	.07	.520	-.14, .26	-.21	.034	-.40, -.01
Amy						
Segment 1	.26	.011	.06, .44	.07	.487	-.13, .27
Segment 2	.02	.812	-.18, .23	.14	.172	-.07, .33
Segment 3	.19	.058	-.01, .38	-.08	.425	-.28, .12
Segment 4	.08	.471	-.14, .28	-.03	.793	-.23, .18
Segment 5	.01	.893	-.19, .22	.05	.596	-.15, .25
Jia						
Segment 1	-.05	.621	-.25, .16	-.13	.198	-.32, .07
Segment 2	.07	.497	-.14, .27	.22	.026	.02, .40
Segment 3	.05	.594	-.15, .25	.06	.546	-.14, .26
Segment 4	-.09	.374	-.29, .12	.04	.679	-.16, .24
Segment 5	.00	.970	-.22, .22	-.02	.863	-.22, .19
Derek						
Segment 1	.07	.499	-.14, .27	.03	.773	-.17, .23
Segment 2	-.11	.264	-.31, .09	-.26	.009	-.44, -.06
Segment 3	-.07	.461	-.27, .13	-.01	.891	-.21, .19
Segment 4	.01	.889	-.19, .22	-.06	.574	-.25, .15

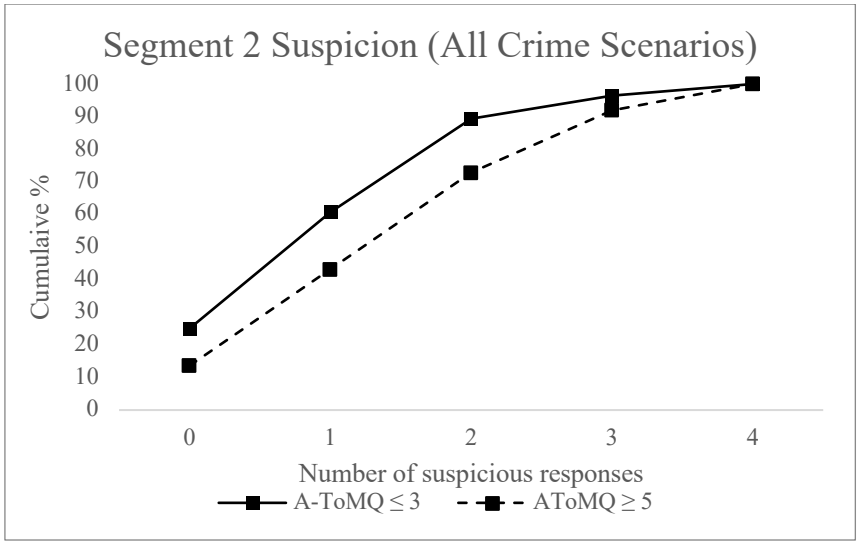
Appendix T

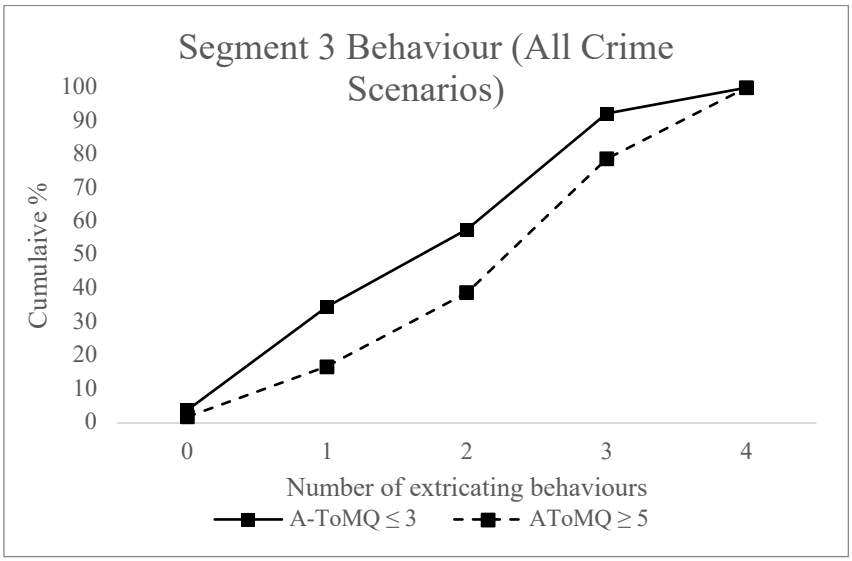
Cumulative percentage graphs representing the number of adaptive behaviours and suspicion responses at each segment for the high and low ToM groups.

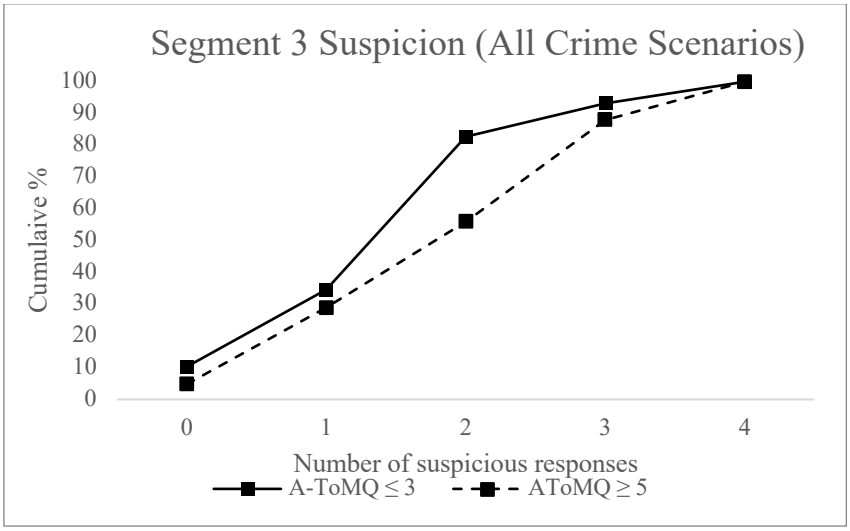


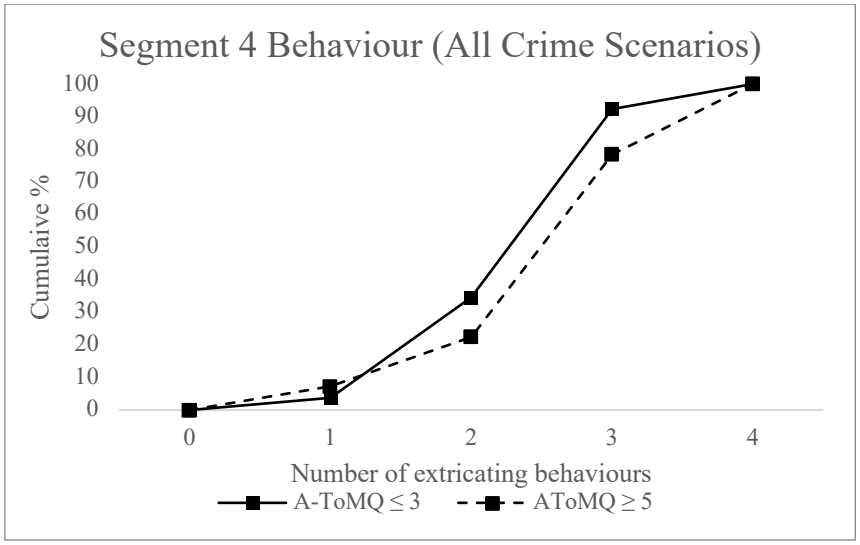


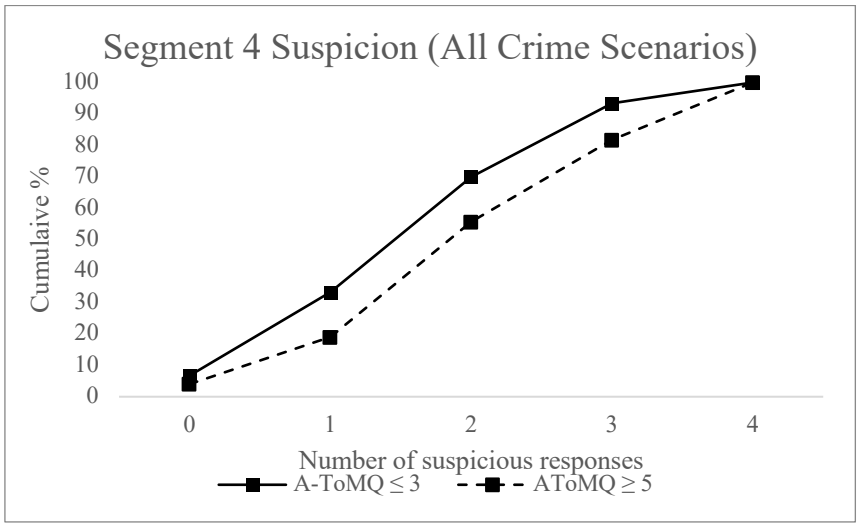



Segment 2 Suspicion (All Crime Scenarios)
Cumulaive %
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
Number of suspicious responses
A-ToMQ ≤ 3
AToMQ ≥ 5


Segment 3 Behaviour (All Crime Scenarios)
Cumulaive %
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
Number of extricating behaviours
A-ToMQ ≤ 3
AToMQ ≥ 5


Segment 3 Suspicion (All Crime Scenarios)
Cumulaive %
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
Number of suspicious responses
A-ToMQ ≤ 3
AToMQ ≥ 5


Segment 4 Behaviour (All Crime Scenarios)
Cumulaive %
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
Number of extricating behaviours
A-ToMQ ≤ 3
AToMQ ≥ 5


Segment 4 Suspicion (All Crime Scenarios)
Cumulaive %
100
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
Number of suspicious responses
A-ToMQ ≤ 3
AToMQ ≥ 5

Appendix U

Descriptive statistics for proposed validity measures in Study 3

	N	Min	Max	Mean	SD	Median	95% CI
Autistic	95						
Gullibility		12	84	38.27	16.62	34.0	35.01, 41.54
SVS		0	37	12.39	7.31	10.0	10.96, 13.83
Trust		43	93	66.25	9.55	68.0	64.38, 68.13
Non-autistic	102						
Gullibility		12	70	30.42	12.75	28.0	27.82, 33.02
SVS		0	37	10.25	5.73	10.0	9.09, 11.42
Trust		44	99	67.71	10.21	67.0	65.63, 69.79

Inferential statistics for proposed validity measures in Study 3

	t	*df*	*p*	*d*	95% CI	*U*	*p*
Gullibility	3.74	188.21	<.001	-0.53	-.81, -.24	6210.5	<.001
SVS	-2.29	189.52	.023	-0.32	-061, -.04	5590.50	.062
Trust	1.03	195	.152	0.15	-0.13, 0.43	4729.0	.772