Original Article

autism

Autism
2023, Vol. 27(4) 1026–1035

DOI: 10.1177/13623613221125564
journals.sagepub.com/home/aut
SAGE

Detecting dodgy behaviour: The role of autism, autistic traits and theory of mind

Neil Brewer, Carmen A Lucas, Alliyza Lim and Robyn L Young

Abstract

We examined whether theory of mind difficulties often considered to characterise autistic individuals impair their ability to detect the presence of dodgy or suspicious behaviour in interactions, thereby rendering them especially vulnerable to becoming involved in criminal activity or being victimised. Using a signal detection theory approach, we compared autistic ($N = 72$; Verbal Comprehension Index = 88–122, $M = 106.6$) and non-autistic ($N = 70$; Verbal Comprehension Index = 86–120, $M = 104.4$) adults' ability to detect dodgy or suspicious behaviour across a broad array of scenarios ($N = 136$). Although theory of mind performance was poorer for the autistic group, frequentist and Bayesian analyses indicated there were no group differences in either the standard measures of discrimination performance obtainable using a signal detection theory approach or in terms of a bias towards reporting dodgy behaviour. Furthermore, there was no indication of a relationship between dodginess detection and autistic traits. However, regardless of group membership, theory of mind difficulties were associated with poorer discrimination of dodgy behaviour, highlighting an individual difference variable that may increase the vulnerability of both autistic and non-autistic individuals to involvement in some form of criminal activity or to becoming a victim of crime.

Lay abstract

Difficulties in reading others' minds make it difficult to anticipate their future behaviour. It has often been argued that such difficulties contribute to autistic individuals becoming enmeshed in criminal activity. However, supportive scientific evidence is virtually non-existent. We compared the ability of groups of autistic and non-autistic adults of similar intellectual ability to detect dodgy or suspicious behaviour across a wide range of scenarios. Although the autistic group performed more poorly than the non-autistic group on an established measure of mindreading, there were no group differences in the ability to detect dodginess. Nor did we find any evidence that detecting dodgy behaviour was associated with the degree of autistic traits reported by individual participants. However, when we combined the two groups, difficulty reading the minds of others was indeed associated with poorer detection of dodginess, thus highlighting a characteristic of individuals that may well increase the likelihood of becoming involved in crime or exploited for autistic and non-autistic individuals alike.

Keywords

autism, autistic traits, criminality, signal detection theory, theory of mind

What factors render individuals vulnerable to becoming involved in crime or being victimised? Predisposing psychopathologies, socio-environmental experiential history, risky social networks and substance and alcohol abuse are among the many factors that have been implicated (e.g. Clarke, 1992; Gottfredson & Hirschi, 1990; Loeber & Dishion, 1983). One individual difference variable that has received some consideration in this context is theory of mind (ToM): namely, the ability to take the perspective or read the mind of other individuals, thereby anticipating their intentions and likely behaviour (Baron-Cohen, 1995,

Flinders University, Australia

Corresponding author:
Neil Brewer, College of Education, Psychology and Social Work, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia.
Email: neil.brewer@flinders.edu.au

2001; Baron-Cohen et al., 1985; Happé, 1994). ToM difficulties might, for example, underpin difficulties recognising the meaning behind an interaction partner's facial expressions, body language and tone of voice; discerning the real meaning underlying verbal expressions involving figurative language; or recognising that socially inappropriate remarks might have an adverse impact on others (e.g. Baron-Cohen, 2001; Brewer & Young, 2015; Happé, 1994). A recent systematic review (Karoğlu et al., 2021) focused on the ToM abilities of criminal offenders and non-offenders and, while highlighting mixed findings and a variety of methodological issues (e.g. sample sizes, cognitive ability controls, psychometric integrity of ToM measures), concluded that impairments in ToM that may underpin problematic interactions with other people appear to be more prevalent in criminal offenders.

Given that it has long been argued that ToM difficulties underpin some of the social difficulties experienced by autistic individuals (e.g. Baron-Cohen, 1995, 2001; Baron-Cohen et al., 1985), it is unsurprising that such difficulties have sometimes been highlighted in clinical case reports and legal arguments focusing on autism-specific influences on involvement in criminal activity (e.g. Baron-Cohen, 1988; Brewer & Young, 2015; Kibbie, 2012; Murrie et al., 2002). Indeed, it does seem plausible that difficulties 'reading' the real intentions of others during social interactions may constrain an individual's ability to recognise subtle cues to the dodgy[1] or suspicious nature of an interaction partner's intentions before it is too late to extricate themself from being lured into criminal activity or being victimised.

However, although pronounced ToM difficulties appear to characterise some proportion of autistic adults, they are not ubiquitous, with substantial variability within autistic (and non-autistic) adult samples (e.g. Brewer et al., 2017, 2022; Gernsbacher & Yergeau, 2019). Moreover, despite supportive case study reports and evidence suggesting a greater likelihood of ToM difficulties among criminal offenders, compelling empirical support for any heightened criminal vulnerability of autistic individuals and for the contribution of ToM difficulties (or indeed other autistic characteristics) to such vulnerability is scarce. Consequently, it is unclear whether ToM difficulties might heighten the vulnerability of autistic individuals in particular or whether such difficulties might contribute to involvement in criminal activity regardless of autism diagnosis or the presence of autistic characteristics. When any individual is at risk of being manipulated or coerced into criminal behaviour by an interaction partner, it is imperative that they recognise the partner's intent and the potential impact of their behaviour on others. Thus, persons with a poorly developed ToM, whether they be autistic or not, may be particularly vulnerable to becoming enmeshed in criminal activity.

Empirical evidence demonstrating that autistic individuals may have difficulty recognising criminal intent, or that ToM may underpin such difficulties, appears to be confined to two published studies. In the first of two experiments, Williams et al. (2018) had a sample of non-autistic college students view videos of liars and truth-tellers and detected a small but significant correlation between a measure of autistic traits and performance on a measure of deception detection. In their second experiment, they reported significant group differences between samples of autistic and non-autistic adults on the same deception detection task. However, no relationships between their ToM measures – the Frith–Happé animations (White et al., 2011) and the Reading the Mind in the Eyes tasks (Baron-Cohen, Wheelwright, Hill et al., 2001) – deception detection performance were detected in Williams et al.'s first experiment and no ToM measure was taken in the second experiment.

The second study (Brewer et al., 2018) examined whether ToM difficulties, measured using the Frith–Happé animations, were related to latency to indicate that there was something suspicious or dodgy about the way in which an interaction was unfolding. Non-autistic participants listened to audio vignettes describing hypothetical scenarios, some of which culminated in criminal activity and some of which did not. During each recording, participants signalled when they detected that something dodgy was unfolding. Brewer et al. detected significant, although relatively weak, negative relationships between ToM and time to react to cues to dodgy behaviour. However, that study only included non-autistic adults; additionally, the latency measures were based on responses to a very small number of vignettes and, given the intra-and inter-individual variability normally associated with latency measures, were likely to be noisy and unreliable estimates of the ability to recognise dodgy behaviour.

Given the absence of strong empirical evidence demonstrating the susceptibility of autistic individuals to being lured into criminal activity or victimised, or for the contributions of ToM difficulties to such outcomes, this study was designed to examine whether (a) autistic adults (and individuals whose AQ scores suggest a higher degree of autistic traits) are impaired in the ability to discriminate cues that should signal dodgy behaviour; (b) difficulty in making such discriminations is associated with ToM difficulties, regardless of an autism diagnosis or the presence of autistic traits; and (c) any autistic–non-autistic differences in discrimination of dodgy behaviour are mediated by ToM. We expected the autistic group to perform more poorly on the ToM measure and hypothesised that (a) autistic adults would be poorer than non-autistic adults at detecting dodgy behaviour, (b) poorer ToM would be associated with poorer discrimination of dodginess regardless of autism diagnosis, and (c) the poorer discrimination of dodginess by autistic adults would be mediated by their particular ToM difficulties.

To examine these questions, we recruited sizable samples of autistic and non-autistic adults and adopted a signal

detection theory approach (Macmillan & Creelman, 2005; Swets, 1988) designed to provide a bias-free measure of the ability to discriminate dodgy or suspicious behaviour across an array of scenarios. Signal detection measures index two aspects of decision-making under uncertainty: discriminability and response bias (Macmillan & Creelman, 2005). Discriminability indicates how effectively the individual distinguishes between the presence or absence of a target signal in a stimulus – in this case, the presence or absence of dodgy behaviour. One index of discriminability is d', calculated as $z(H) - z(F)$ where H is the hit rate (i.e. the number of correct dodgy or 'suspicious' responses divided by the number of dodgy trials), F is the false alarm rate (i.e. the number of incorrect 'dodgy' responses divided by the number of non-dodgy trials) and z is the inverse of the standard normal cumulative distribution (also known as the quantile function). A d' value of 0 indicates chance performance and increasingly higher values indicate increasing levels of discriminability. Response bias indicates whether the decision maker is inclined towards saying a signal is present or absent (i.e. regardless of whether the signal is present or not) and is commonly measured by criterion placement (or c). Criterion placement is calculated as $-0.5\ [z(H) + z(F)]$. A c value of 0 indicates unbiased decision-making – in this case neither a preference for responding 'dodgy' nor 'not dodgy'. Values below 0 indicate a bias towards reporting signal (i.e. responding 'dodgy') and values above 0 indicate a bias towards not reporting signal (i.e. responding 'not dodgy').

Receiver operating characteristics (ROC) analysis provides another index of discrimination performance in situations like the present context where the perceived suspiciousness or dodginess of a scenario was also measured on a scale from 'definitely not dodgy' to 'definitely dodgy' (Swets, 1988). ROC analysis involves plotting the hit rate against the false alarm rate (i.e. sensitivity against 1–specificity) across the range of possible cut-off scores, resulting in a visual depiction of discriminability. The level of discrimination performance is indicated by how far the resultant curve bends towards the top left-hand corner of the graph. The associated area under the curve (AUC) statistic, ranging from 0.5 (chance performance) to 1 (perfect discrimination), provides a numerical index of discrimination performance.

These signal detection theory statistics provide the basis for an examination of the hypotheses outlined previously. Participants read 136 short-written vignettes that were extracted from short stories, some of which culminated in criminal activity while the others did not. The vignettes were pre-tested to ensure they captured a range of signal strengths to dodginess: for example, some strongly signalled dodgy behaviour, some weakly signalled dodgy behaviour. After each vignette, participants indicated whether or not something dodgy was going on and how certain they were about that judgement. ToM was assessed using the social sub-scale of the Adult Theory of Mind-Quick test (A-ToM-Q; Brewer et al., 2022).

Method

Participants

To detect a medium-sized difference in discrimination performance between groups ($\alpha = 0.05$, power = 0.80, tail(s) = two), our target sample size was 128 (64 per group). Based on a pilot study's estimates of likely numbers who would fail attention checks or not return for the required second experimental session, we targeted 78 participants for each group. One hundred and fifty-eight participants were recruited. Five did not complete session 2, and 11 others were excluded for failing attention or vocabulary checks or scoring below 85 on a verbal IQ measure.

The final autistic sample comprised 72 adults (33 male, 34 female, 5 non-binary), aged 19–64 years (M = 33.0, standard deviation (SD) = 10.9, Mdn = 30.0), drawn from autistic individuals registered on the Prolific crowdsourcing platform. Verbal ability was estimated using the Spot-the-Word Task (STW; Baddeley et al., 1993), which has been rigorously evaluated (e.g. Mackinnon & Christensen, 2007), demonstrates convergent validity with the Wechsler Adult Intelligence Scale (Yuspeh & Vanderploeg, 2000), and can be used to predict WAIS-IV Verbal Composite Index (VCI; Baddeley & Crawford, 2012). Estimated VCI based on this instrument ranged from 88 to 122 (M = 106.6, SD = 8.2, Mdn = 107.0). Autism diagnosis was ascertained by asking participants to provide detailed information about their diagnosis, including confirmation of a formal diagnosis by a qualified clinical professional, age at diagnosis, diagnosis date and location, and the diagnostician. Autistic participants reported having been diagnosed between the ages of 4 and 59 (M = 22.0, SD = 12.4, Mdn = 21.5), usually by a psychiatrist ($N = 41$) or psychologist ($N = 21$); other sources of diagnosis included family doctors ($N = 5$) or an autism specialist ($N = 4$) and the origin of one childhood diagnosis was reported as unknown. The non-autistic sample comprised 70 adults (29 male, 39 female, 1 non-binary, 1 preferred not to specify gender), aged 18–73 years (M = 38.9, SD = 14.0, Mdn = 37.0), with estimated VCI scores ranging from 86 to 120 (M = 104.4, SD = 7.4, Mdn = 104.0).

The autistic group scored significantly higher on a measure of autistic characteristics, the AQ-12 (Lundqvist & Lindner, 2017; $M = 10.18$, SD = 2.79, Mdn = 11.00), than the non-autistic group (M = 5.26, SD = 3.38, Mdn = 5.00), $t(133.55) = 9.46$, $p < 0.001$, $d = 1.59$, 95% confidence interval (CI): [1.21, 1.97], $U = 7.86$, $p < 0.001$. There was no significant difference in the proportions of male and female participants, $\chi^2(1) = 0.59$, $p = 0.441$, or in estimated VCI, $t(140) = 1.68$, $p = 0.095$, $U = 1.71$, $p = 0.087$. The non-autistic sample was, however, older than the autistic

sample, $t(130.12) = 2.78$, $p = 0.006$, $d = 0.47$, 95% CI: [0.14, 0.80], $U = 2.48$, $p = 0.013$.

Measures

Discrimination of dodgy behaviour. The stimulus material comprised 136 written vignettes, ranging in length from 45 to 267 words. The vignettes were segments extracted from 34 professionally short-written stories that ranged in length from 513 to 870 words. Twenty of the 34 short stories culminated in a crime being committed; the others did not. Eighty-one vignettes were from crime stories and 55 were from non-crime stories. One example of each type of story appears in Supplementary Material Table S1; the full set can be accessed at https://osf.io/rn8xv/?view_only=1136e8efd76248bbb089c1d9bb17d067.

Initially, the stories were segmented into 152 brief vignettes that were judged by two of the authors as (a) able to 'stand-alone' and (b) likely to be rated as varying widely in the likelihood that they would be judged as suggestive of dodgy behaviour. The objective was to identify sub-sets of vignettes that would be almost universally regarded (e.g. by 90%–100% of independent observers) as suggestive of dodgy behaviour, regarded by progressively smaller proportions of observers (e.g. 70%–90%, 50%–70%, 30%–50%, 10%–30%) as suggestive of dodgy behaviour, and almost universally regarded as not suggestive of dodgy behaviour.

To ensure the resulting vignettes encompassed a broad range of signal strength to dodgy behaviour, a pilot study had MTurk observers ($N = 152$) read all vignettes and (a) classify (Yes–No) whether each vignette indicated dodgy or suspicious behaviour or not, and (b) provide a confidence judgement to indicate the presence of dodgy behaviour as 0 = *Sure Not Dodgy*, 20 = *Probably Not Dodgy*, 40 = *Maybe Not Dodgy*, 60 = *Maybe Dodgy*, 80 = *Probably Dodgy*, 100 = *Sure Dodgy.* Supplementary Material Table S2 shows (a) the number of vignettes rated dodgy by different percentages of participants, and (b) the number of vignettes assigned to each of the different confidence ratings. The data indicate that, as intended, the strength of the cues to dodginess in the vignettes varied relatively evenly from very weak to very strong.

To reduce the number of vignettes to a more manageable number for participants to complete (along with our other measures) in two online test sessions, we removed 16 vignettes: 8 from crime stories that very few observers rated as dodgy and 8 from non-crime scenarios that many observers rated as dodgy. Supplementary Material Table S3 provides two examples of vignettes considered suggestive of suspicious behaviour by 91%–100%, 51%–70% and 0%–10% of observers. The full set can be accessed at https://osf.io/rn8xv/?view_only=1136e8efd76248bbb089c1d9bb17d067.

The 136 vignettes (plus 6 vignettes used as attention checks) were presented in random order across two separate testing sessions, with participants receiving the following instructions:

- You are going to be shown a large number of brief scenarios. After you have read each one, you will be asked to respond to two questions.
- You will be asked to indicate if you think there is 'anything dodgy' going on in the scenario by clicking on *Yes* or *No*. *Something 'dodgy' means that there is something suspicious about the situation or the way in which someone is behaving.*
- You will then be asked 'how sure are you?' – you should respond by clicking on one of the six buttons to indicate that you are sure it is dodgy, probably dodgy, maybe dodgy, maybe not dodgy, probably not dodgy or sure it is not dodgy.
- At that point you should click the arrow at the bottom of the page to proceed to the next scenario. There will be 71 scenarios in each session.

Please note:

- In some scenarios it may seem fairly obvious that something dodgy is going on. In others it may seem fairly obvious that nothing dodgy is happening. And in others it may not be so clear either way.
- Some scenarios may appear to be related to others you encountered previously, but please consider each scenario independently when making your response.
- Try to give your first impression rather than stewing over your answer.
- There is no right or wrong answer.

ToM. ToM was measured with the six-item social scale of the A-ToM-Q test (Brewer et al., 2022). The A-ToM-Q uses the same stimulus items as the A-ToM (Brewer et al., 2017) but requires multiple-choice instead of free-report responding. The social, or perspective taking, items require examinees to interpret the real meaning of elements such as sarcasm, faux pas, white lie, bluff, persuasion and misunderstanding that are depicted in brief videos of real-life interactions. Both the A-ToM and the A-ToM-Q have been subjected to comprehensive psychometric examinations which have highlighted the instrument's reliability, concurrent, convergent, divergent and discriminant validity (Brewer et al., 2017, 2022). Importantly, the A-ToM-Q's social scale differentiated autistic and non-autistic adults much more strongly than its physical or control scale which did not require social or perspective taking inferences.

The ToM scale items can be viewed at the URL below; the scoring protocol is described in Brewer et al. (2022):

https://www.youtube.com/playlist?list=PLJCW1evzKKctzHvYfB1RADd27m8IBaWcu

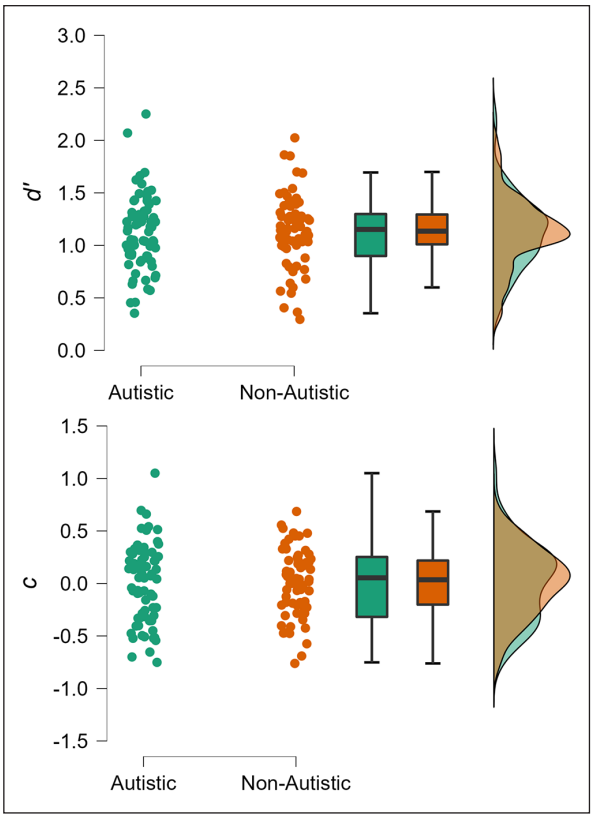

Figure 1. Raincloud plots showing the distributions of *d'* (top panel) and *c* (bottom panel) in the autistic and non-autistic groups.

Items were presented in random order with each followed by four randomly ordered forced-choice response options, with participants asked to respond as quickly and accurately as possible (see Brewer et al., 2022). Responses were scored 0 (incorrect) or 1 (correct), with possible total scores ranging from 0 to 6. Item response latency for each item was recorded from the appearance of the multiple-choice alternatives to the participant's mouse click on their chosen alternative.

Verbal ability. The STW task (Baddeley et al., 1993) was used to estimate VIQ. Participants view 100 word-pairs, each comprising a real word (e.g. albatross) and non-word (e.g. zando) and are required to identify the real word. Their accuracy scores can be used to estimate WAIS-IV VCI (Baddeley & Crawford, 2012). The test has been carefully evaluated, demonstrating strong convergent validity with other measures of vocabulary and fluid intelligence (e.g. Baddeley et al., 1993; Mackinnon & Christensen, 2007; Yuspeh & Vanderploeg, 2000).

Autistic characteristics. The AQ-12 (Lundqvist & Lindner, 2017) was used to measure the presence of autistic traits ($\alpha = 0.80$ for our sample). It is a 12-item brief version of the widely used Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner et al., 2001), which is a 50-item self-report screening measure of autism spectrum disorder (ASD), designed for adults of normal intelligence. AQ scores reflect the notion that autistic traits can be represented as a continuum, with higher scores indicative of a higher degree of autistic characteristics. Lundqvist and Lindner (2017) showed that the AQ could be reduced to a 12-item measure with little loss of explanatory power. This allowed us to minimise time demands on our participants, which was an important consideration given the long duration of the signal detection task sessions.

Procedure

The project was approved by the Human Ethics Low Risk Panel of Flinders University; participants read a study information sheet and gave informed consent. Participants confirmed that they were using a laptop device or similar, answered two questions designed to exclude bots, and responded to screening questions ensuring normal or corrected to normal vision and hearing. They then provided details of their age, gender, native language, and autism diagnosis. Participants were told the tasks would take approximately 75 min to complete in each of two sessions and that they could take breaks during the sessions. The AQ-12 and the A-ToM-Q social scale were completed in session 1 and the STW test in session 2; the signal detection vignettes were administered over two sessions. Participants received an honorarium in recognition of their significant time commitment to the project.

Community involvement statement

Two of the authors are practicing clinical psychologists who consult with autistic adults and children.

Results

Our analyses focus on group differences in dodginess discriminability (and response bias), the relationship between autistic characteristics and discriminability and the relationship between ToM and discriminability.

Group differences in discrimination and bias

Signal detection measures of discriminability (d') and response bias (c) were calculated from the binary (dodgy–not dodgy) responses to the vignettes.[2] As shown in Figure 1, the distributions of both d' and c were very similar for the autistic and non-autistic groups.

For the autistic group, d' ranged from 0.35 to 2.25 ($M = 1.12$, $SD = 0.35$, $Mdn = 1.15$) and c ranged from −0.75 to 1.05 ($M = -0.01$, $SD = 0.37$, $Mdn = 0.05$). For the non-autistic group, d'' ranged from 0.30 to 2.02 ($M = 1.13$, $SD = 0.34$, $Mdn = 1.14$) and c ranged from −0.76 to 0.69

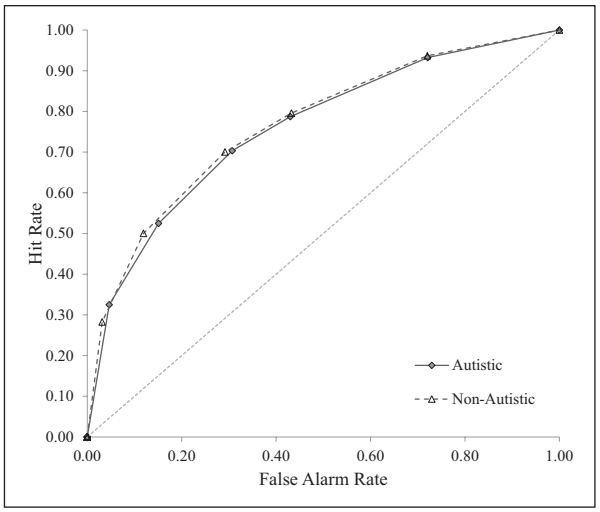

Figure 2. ROC curves showing criterion-free discriminability between dodgy and non-dodgy scenarios for the autistic and non-autistic groups.

(M = 0.01, SD = 0.31, Mdn = 0.04). There were no significant differences in d' or c between the two groups, $t(140) = 0.21$, $p = 0.834$, $d = 0.03$, 95% CI: [−0.30, 0.36], $U = 0.40$, $p = 0.692$, and $t(140) = 0.33$, $p = 0.745$, $d = 0.06$, 95% CI: [−0.27, 0.39], $U = 0.38$, $p = 0.704$, respectively.

The dodginess confidence ratings (i.e. maybe dodgy/not dodgy, probably dodgy/not dodgy, sure dodgy/not dodgy) were used to construct ROC curves (see Figure 2) across all trials for the autistic and non-autistic groups. AUC at the trial level was 0.759, 95% CI: [0.750, 0.768] for the autistic group and 0.768, 95% [0.759, 0.777] for the non-autistic group. AUC was also calculated separately for each participant. For the autistic group, AUC scores ranged from 0.589 to 0.900 (M = 0.767, SD = 0.062, Mdn = 0.781), and did not differ meaningfully from those of the non-autistic group, $t(140) = 0.59$, $p = 0.557$, $d = 0.10$, 95% CI: [−0.23, 0.43], $U = 0.43$, $p = 0.665$, with the latter's AUC scores ranging from 0.614 to 0.896 (M = 0.773, SD = 0.058, Mdn = 0.780).

Bayesian analyses of the effects of group on d', AUC and c provided moderate evidence for the null hypothesis in each case. The respective non-parametric Bayesian statistics (calculated from 10,000 bootstrap samples) for d', AUC and c were $BF_{10} = 0.186$, $W = 2423.00$, $Rhat = 1.00$; $BF_{10} = 0.192$, $W = 2414.00$, $Rhat = 1.00$; and $BF_{10} = 0.188$, $W = 2427.00$, $Rhat = 1.00$.

The relationship between autistic characteristics and discriminability

AQ-12 scores were not correlated with d', $r_s(140) = 0.085$, $p = 0.315$; AUC, $r_s(140) = 0.070$, $p = 0.411$; or c, $r_s(140) = -0.094$, $p = 0.264$. These negligible relationships were replicated when correlating d' with the AQ-12's social and communication items, $r_s(140) = 0.061$, $p = 0.471$, and attention switching items, $r_s(140) = 0.084$, $p = 0.319$. We also examined the possibility that there might be some AQ-12 threshold level above which poorer discriminative performance might be detected. Table 1 shows the comparison of d' and AUC scores using each possible AQ-12 score as the threshold or cut-off: that is, comparing d' and AUC for participants scoring 0 versus ⩾1; ⩽1 versus ⩾2; ⩽2 versus ⩾3; . . . through to ⩽11 versus 12. Regardless of the threshold AQ level examined, there were no significant differences in d' or AUC.

The relationship between ToM and discrimination of dodginess

As previously found with samples of autistic and non-autistic adults for the A-ToM-Q (Brewer et al., 2022) and the A-ToM (Brewer et al., 2017), there was considerable inter-individual variability in the ToM scores within each group and substantial overlap between groups. However, as expected, the autistic group (M = 4.21, SD = 1.65, Mdn = 5.00) scored significantly lower on the A-ToM-Q social scale than the non-autistic group (M = 5.31, SD = 0.88, Mdn = 6.00), $t(108.72) = 5.00$, $p < 0.001$, $d = 0.83$, 95% CI: [0.49, 1.17], $U = 4.36$, $p < 0.001$, with the pattern replicated across individual items (see Supplementary Material Table S4). As shown in Table 2, only autistic participants (a subset comprising 15.3% of the group) scored lower than 3 out of 6 on the A-ToM social scale.

Higher levels of ToM were associated with stronger discrimination of dodginess as indexed by both d' and AUC (see Table 2), with this pattern reflected in significant correlations between ToM and both d', $r_s(140) = 0.295$, $p < 0.001$, and AUC, $r_s(140) = 0.323$, $p < 0.001$.[3] ToM was not, however, correlated with response bias (c), $r_s(140) = -0.098$, $p = 0.248$. In the absence of any group differences in discriminability, the hypothesised mediation relationship was not conducted.

Although no correlation was detected between IQ and ToM, $r_s(140) = 0.033$, $p = 0.969$, we also examined the partial correlations between ToM and the discrimination indices with IQ controlled, given the presence of significant correlations between IQ and both d', $r_s(140) = 0.344$, $p < 0.001$, and AUC, $r_s(140) = 0.373$, $p < 0.001$. The significant correlations between ToM and both d', $r_{partial}(139) = 0.302$, $p < 0.001$, and AUC, $r_{partial}(139) = 0.336$, $p < 0.001$, were undiminished. The nonsignificant correlation between ToM and c, $r_{partial}(139) = -0.097$, $p = 0.255$, was also unaffected.

Discussion

We examined whether ToM difficulties considered to characterise autistic individuals impede the ability to detect dodgy or suspicious behaviour that might render them vulnerable to

Table 1. Comparisons of *d'* and AUC for low and high levels of autistic traits using each possible AQ-12 threshold.

AQ threshold	*d'*								Difference
	Low level of autistic traits				High level of autistic traits				
	N	Mean (SD)	Range	Median	*N*	Mean (SD)	Range	Median	
0 vs ⩾1	5	1.20 (0.35)	0.68–1.54	1.24	137	1.12 (0.34)	0.30–2.25	1.14	$U = 0.78$, $p = 0.435$
⩽1 vs ⩾2	15	1.06 (0.45)	0.30–2.02	1.06	127	1.13 (0.33)	0.35–2.25	1.14	$U = 0.65$, $p = 0.513$
⩽2 vs ⩾3	23	1.09 (0.40)	0.30–2.02	1.08	119	1.13 (0.33)	0.35–2.25	1.14	$U = 0.22$, $p = 0.825$
⩽3 vs ⩾4	29	1.06 (0.38)	0.30–2.02	1.08	113	1.14 (0.33)	0.35–2.25	1.14	$U = 0.74$, $p = 0.457$
⩽4 vs ⩾5	33	1.03 (0.39)	0.30–2.02	1.07	109	1.16 (0.32)	0.35–2.25	1.14	$U = 1.36$, $p = 0.173$
⩽5 vs ⩾6	44	1.06 (0.36)	0.30–2.02	1.07	98	1.16 (0.33)	0.35–2.25	1.16	$U = 1.24$, $p = 0.217$
⩽6 vs ⩾7	53	1.07 (0.34)	0.30–2.02	1.08	89	1.16 (0.34)	0.35–2.25	1.17	$U = 1.34$, $p = 0.179$
⩽7 vs ⩾8	59	1.07 (0.34)	0.30–2.02	1.07	83	1.17 (0.34)	0.35–2.25	1.18	$U = 1.60$, $p = 0.109$
⩽8 vs ⩾9	63	1.09 (0.35)	0.30–2.02	1.11	79	1.15 (0.34)	0.35–2.25	1.18	$U = 0.99$, $p = 0.321$
⩽9 vs ⩾10	80	1.11 (0.33)	0.30–2.02	1.12	62	1.15 (0.35)	0.35–2.25	1.18	$U = 0.65$, $p = 0.518$
⩽10 vs ⩾11	90	1.11 (0.33)	0.30–2.02	1.12	52	1.15 (0.37)	0.35–2.25	1.18	$U = 0.56$, $p = 0.579$
⩽11 vs 12	109	1.11 (0.32)	0.30–2.02	1.13	33	1.17 (0.40)	0.45–2.25	1.20	$U = 0.54$, $p = 0.587$
	AUC								
0 vs ⩾1	5	0.805 (0.060)	0.725–0.882	0.792	137	0.769 (0.060)	0.589–0.900	0.779	$U = 1.18$, $p = 0.236$
⩽1 vs ⩾2	15	0.760 (0.080)	0.614–0.882	0.768	127	0.771 (0.057)	0.589–0.900	0.781	$U = 0.45$, $p = 0.654$
⩽2 vs ⩾3	23	0.766 (0.067)	0.614–0.882	0.781	119	0.771 (0.058)	0.589–0.900	0.781	$U = 0.24$, $p = 0.808$
⩽3 vs ⩾4	29	0.762 (0.062)	0.614–0.882	0.768	113	0.772 (0.059)	0.589–0.900	0.781	$U = 0.95$, $p = 0.340$
⩽4 vs ⩾5	33	0.754 (0.065)	0.614–0.882	0.768	109	0.775 (0.057)	0.589–0.900	0.782	$U = 1.63$, $p = 0.092$
⩽5 vs ⩾6	44	0.764 (0.062)	0.614–0.882	0.778	98	0.773 (0.059)	0.589–0.900	0.781	$U = 0.69$, $p = 0.489$
⩽6 vs ⩾7	53	0.762 (0.060)	0.614–0.882	0.771	89	0.775 (0.059)	0.589–0.900	0.784	$U = 1.34$, $p = 0.180$
⩽7 vs ⩾8	59	0.764 (0.061)	0.614–0.896	0.777	83	0.775 (0.058)	0.589–0.900	0.784	$U = 1.22$, $p = 0.222$
⩽8 vs ⩾9	63	0.767 (0.061)	0.614–0.896	0.777	79	0.773 (0.059)	0.589–0.900	0.782	$U = 0.72$, $p = 0.471$
⩽9 vs ⩾10	80	0.769 (0.061)	0.614–0.900	0.778	62	0.772 (0.059)	0.589–0.898	0.783	$U = 0.65$, $p = 0.516$
⩽10 vs ⩾11	90	0.769 (0.058)	0.614–0.900	0.778	52	0.772 (0.063)	0.589–0.898	0.783	$U = 0.62$, $p = 0.535$
⩽11 vs 12	109	0.770 (0.060)	0.589–0.900	0.778	33	0.772 (0.061)	0.644–0.898	0.782	$U = 0.23$, $p = 0.817$

SD: standard deviation.

Table 2. *d'* and AUC scores for the autistic and non-autistic groups at each level of the A-ToM-Q.

A-ToM-Q score	*d'*							
	Autistic				Non-autistic			
	N	Mean (SD)	Range	Median	*N*	Mean (SD)	Range	Median
0	2	0.64 (0.40)	0.35–0.92	0.64	0	–	–	–
1	5	1.03 (0.30)	0.71–1.51	1.01	0	–	–	–
2	4	1.17 (0.60)	0.84–2.04	0.88	0	–	–	–
3	10	1.09 (0.37)	0.63–1.62	1.15	4	0.81 (0.31)	0.41–1.06	0.89
4	14	1.10 (0.22)	0.57–1.43	1.17	7	0.81 (0.37)	0.30–1.39	0.68
5	18	1.16 (0.44)	0.45–2.25	1.17	22	1.11 (0.31)	0.54–2.02	1.15
6	19	1.18 (0.27)	0.58–1.66	1.23	37	1.24 (0.29)	0.36–1.86	1.23
	AUC							
0	2	0.660 (0.100)	0.589–0.730	0.660	0	–	–	–
1	5	0.751 (0.053)	0.688–0.833	0.741	0	–	–	–
2	4	0.737 (0.060)	0.671–0.813	0.731	0	–	–	–
3	10	0.750 (0.072)	0.644–0.859	0.752	4	0.689 (0.047)	0.640–0.739	0.689
4	14	0.778 (0.036)	0.701–0.820	0.788	7	0.724 (0.068)	0.614–0.820	0.725
5	18	0.776 (0.067)	0.644–0.898	0.786	22	0.769 (0.054)	0.646–0.882	0.773
6	19	0.782 (0.057)	0.687–0.900	0.782	37	0.794 (0.045)	0.642–0.896	0.792

SD: standard deviation.

involvement in criminal activity or victimisation. Although the autistic group scored significantly lower than the non-autistic group on the ToM measure, the indices of discriminability did not demonstrate the hypothesised group differences in the detection of dodgy behaviour. Nor was there a significant relationship between the presence of autistic traits and dodginess detection. There was, however, a significant and moderate relationship between ToM and dodginess detection. Thus, ToM difficulties – rather than autism diagnosis or the presence of autistic characteristics – were associated with individuals' ability to detect the presence of dodgy behaviour.

These findings do not necessarily mean that inferences drawn from case studies of autistic individuals' involvement in crime are incorrect. Perhaps, however, when an offender has a known autism diagnosis, speculation about factors contributing to their involvement in criminal activity may be more likely to encompass consideration of ToM difficulties than might be the case when speculating about non-autistic individual's offending. And almost certainly there will be other factors contributing to detection of dodgy behaviour that do not render it inevitable that an autistic individual scoring below a certain point on a ToM scale will fail to detect dodginess. Detecting dodginess is likely to involve more than simply recognising other people's intentions or motivations, especially as that will not always be possible, regardless of one's ToM. There may be other autistic characteristics that, given the appropriate environmental conditions, interact with ToM and reduce the likelihood of individuals detecting and responding to cues to dodgy behaviour. For example, it has been argued that if a situation presents an opportunity to make some desired social connection or engage in an intense interest, the potential negative impact of an autistic individual's ToM difficulties may be more likely realised (see, for example, many of the case studies described by Brewer and Young (2015)). Moreover, although we are not aware of other empirical research that has isolated variables underpinning dodginess detection, it seems plausible that factors such as IQ, and prior experiences and acquired knowledge, that have helped the acquisition of 'street-smarts' could enhance the individual's sensitivity to suspicious contextual or environmental cues. Although measuring IQ might be relatively straightforward, operationalising and assessing variables such as 'street-smarts' and the way they interact with ToM are likely to be much trickier.

We also must emphasise, as our data indicate, that the likelihood of identifying individuals with pronounced ToM difficulties appears to be higher in autistic samples. Thus, ToM difficulties may be very important contributors to criminal involvement and victimisation for a very small subset of autistic individuals, a risk factor of which justice system professionals should be cognisant. Note also that (as foreshadowed above), ToM difficulties may emerge as more influential for lower functioning autistic individuals than those in our sample. In this context, we reiterate the relationship we observed between IQ and dodginess detection, a relationship long recognised in research on victimisation of individuals with intellectual disabilities (e.g. Wilson & Brewer, 1992).

We also think it is important to link this study with recent research indicating that, when suspected of being involved in suspicious activities, autistic adults may be less effective than non-autistic peers at identifying those elements of the situation that would allay suspicions about their involvement, with those group differences mediated by ToM (Young & Brewer, 2020). Thus, although ToM difficulties may not render autistic individuals less likely to detect dodginess than non-autistic individuals, once they are involved in a problematic situation their ToM difficulties may limit the ability to allay suspicions about their involvement.

Finally, we emphasise limitations of our study which point to various directions for future research. First, the participant's challenge in our dodginess detection paradigm is obviously different from detecting criminal intent in a real-life interaction, intent that may be cued by an array of often subtle visual and language cues embedded within complex and dynamic environmental stimuli. Second, precisely how such contextual differences might shape group differences in dodginess discrimination is unknown. Third, there is a need to investigate the links between discrimination of cues to dodgy behaviour and subsequent involvement in criminal activity. Fourth, despite the negligible relationship between dodginess discrimination and the AQ-12, we note that the AQ-12 items do not capture the presence of repetitive behaviours and intense interests. Consequently, as already noted, we cannot rule out the possibility that such characteristics may, under some conditions, be related to dodginess detection. Fifth, as emphasised by Karoğlu et al. (2021), it will be important to distinguish the roles of cognitive and affective components of ToM, both of which are likely captured to some degree in some items of our ToM measure. Sixth, the relationships between the indices of dodginess discrimination and ToM are, of course, correlational in nature and thus our research must be seen as a starting point for the systematic investigation of these issues.

Although follow-up of issues such as those just outlined is imperative, we conclude by emphasising that our findings pinpoint an individual difference variable that may heighten the vulnerability of some individuals – autistic and non-autistic – to becoming ensnared as participants or victims of some form of criminal activity.

Acknowledgements

We thank Lauren Fuge who authored the short stories from which our stimulus materials were drawn.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Research supported by ARC DP 190100162 and the Hamish Ramsay Fund.

ORCID iDs

Neil Brewer https://orcid.org/0000-0003-3465-9244
Carmen A Lucas https://orcid.org/0000-0002-7984-8068
Alliyza Lim https://orcid.org/0000-0001-7451-6112
Robyn L Young https://orcid.org/0000-0002-6124-8515

Data availability

The data, analysis code, and materials are available on the Open Science Framework at https://osf.io/rn8xv/?view_only=1136e8efd76248bbb089c1d9bb17d067

Supplemental material

Supplemental material for this article is available online.

Notes

1. 'Dodgy' is a familiar term in countries such as the United Kingdom and Australasia that refers to suspicious, unreliable, dishonest actions or persons.
2. For 1 of the 55 non-dodgy vignettes, an unidentified software fault resulted in missing responses from 27.9% of participants; all calculations excluded this trial.
3. Median latency for responses to the six theory of mind (ToM) items did not correlate significantly with d', $r_s(141) = 0.127$, $p = 0.131$.

References

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

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. https://doi.org/10.1111/j.1469-7610.1988.tb00723.x

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press/Bradford.

Baron-Cohen, S. (2001). Theory of mind and autism: A review. *Special Issue of the International Review of Mental Retardation*, *23*, 169–184. https://doi.org/10.1016/S0074-7750(00)80010-5

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'?. *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'reading the mind in the eyes' test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

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/0.1023/A:1005653411471

Brewer, N., Bay Wei Ying, A., Young, R. L., & Nah, Y.-H. (2018). Theory of mind and the detection of suspicious behavior. *Journal of Applied Research in Memory and Cognition*, *7*(1), 123–131. https://doi.org/10.1037/h0101817

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

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/s10803-017-3150-0

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

Clarke, R. (1992). *Situational crime prevention*. Harrow and Heston.

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

Gottfredson, M., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.

Happé, F. G. E. (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*, 129–154. https://doi.org/10.1007/BF02172093

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

Kibbie, K. S. (2012). Maleficent or mindblind: Questioning the role of Asperger's in quant hedge fund malfeasance and modeling disasters. *The American Criminal Law Review*, *49*(2), 367–402.

Loeber, R., & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin*, *94*(1), 68–99. https://doi.org/10.1037/0033-2909.94.1.68

Lundqvist, L., & 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), 459–468. https://doi.org/10.1037/1040-3590.19.4.459

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. *Psychology Press*.

Murrie, D. C., Warren, J. I., Kristiansson, M., & Dietz, P. E. (2002). Asperger's syndrome in forensic settings. *The*

International Journal of Forensic Mental Health, *1*(1), 59–70. https://doi.org/10.1080/14999013.2002.10471161

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of theory of mind for adults with autism. *Autism Research*, *4*(2), 149–154. https://doi.org/10.1002/aur.174

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 victimization of individuals with an intellectual disability. *Australian Psychologist*, *27*(2), 114–117. https://doi.org/10.1080/00050069208257591

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

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