Metacognitive monitoring and the hypercorrection effect in autism and the general population: Relation to autism(-like) traits and mindreading

David M Williams[1], Zara Bergström[1] and Catherine Grainger[2]

Abstract

Among neurotypical adults, errors made with high confidence (i.e. errors a person strongly believed they would not make) are corrected more reliably than errors made with low confidence. This 'hypercorrection effect' is thought to result from enhanced attention to information that reflects a 'metacognitive mismatch' between one's beliefs and reality. In Experiment 1, we employed a standard measure of this effect. Participants answered general knowledge questions and provided confidence judgements about how likely each answer was to be correct, after which feedback was given. Finally, participants were retested on all questions answered *in*correctly during the initial phase. Mindreading ability and autism spectrum disorder–like traits were measured. We found that a representative sample of ($n = 83$) neurotypical participants made accurate confidence judgements (reflecting good metacognition) and showed the hypercorrection effect. Mindreading ability was associated with autism spectrum disorder–like traits and metacognition. However, the hypercorrection effect was non-significantly associated with mindreading or autism spectrum disorder–like traits. In Experiment 2, 11 children with autism spectrum disorder and 11 matched comparison participants completed the hypercorrection task. Although autism spectrum disorder children showed significantly diminished metacognitive ability, they showed an undiminished hypercorrection effect. The evidence in favour of an undiminished hypercorrection effect (null result) was moderate, according to Bayesian analysis (Bayes factor = 0.21).

Keywords

autism, hypercorrection, mentalising, metacognition, mindreading

Autism spectrum disorder (ASD) is a developmental disorder diagnosed on the basis of severe impairments in social communication, as well as a restricted and repetitive repertoire of behaviour and interests (American Psychiatric Association (APA), 2013). At the cognitive level, ASD is characterised by a diminished ability to attribute mental states to others in order to explain and predict their behaviour (henceforth termed mindreading; also known as theory of mind). Moreover, the extent of mindreading ability/impairment has been shown in some studies to predict the severity of ASD features in those diagnosed with the disorder (see Brunsdon and Happé, 2014) and the number of ASD-like traits in the general population (e.g. Baron-Cohen et al., 2001b). Despite extensive research on mindreading in ASD over the past three decades, very little research has investigated the extent to which individuals with this disorder are able to represent their *own* mental states accurately (so-called metacognition or metacognitive monitoring; see Williams, 2010). This is surprising given many researchers

suggest metacognitive monitoring is thought to rely on the same metarepresentational ability as mindreading (e.g. Carruthers, 2009; Frith and Happé, 1999), so might be expected to be as impaired as mindreading in ASD and equally predictive of ASD features. Furthermore, metacognitive monitoring ability is known to be involved in self-regulated learning and to predict learning/educational outcomes independent of general intelligence in the neurotypical population (e.g. Hartwig and Dunlosky, 2012; Veenman et al., 2005). Thus, a diminution of metacognitive monitoring ability in ASD might well contribute to

[1]University of Kent, UK
[2]University of Stirling, UK

Corresponding author:
David M Williams, School of Psychology, Keynes College, University of Kent, Canterbury CT2 7NP, UK.
Email: D.M.Williams@kent.ac.uk

well-documented educational underachievement that is seen even among intellectually able individuals with ASD (e.g. Estes et al., 2011; Jones et al., 2009).

What relatively little research has been conducted on metacognitive monitoring accuracy in ASD appears to indicate a deficit, although not conclusively so. Children with ASD have well-established difficulties with 'self-versions' of classic mindreading tasks (see Williams, 2010), as well as in real-world learning situations (Brosnan et al., 2016), that require the attribution of mental states to self in order to explain one's own behaviour. Furthermore, adults with ASD report high levels of alexithymia, a difficulty identifying and labelling one's own emotional states (e.g. Griffin et al., 2015). However, it is arguable whether or not these paradigms really measure the accuracy with which one can monitor *online* one's own *current* mental states (see Carruthers, 2009; but see Williams, 2010). Recently, studies of ASD have begun to employ classic paradigms from the field of metacognition that require participants to make explicit online judgements about the state of their own cognition. In these paradigms, participants perform a cognitive task (e.g. answering general knowledge questions) and are asked to make judgements about the accuracy of their responses/knowledge. 'Object-level' performance (i.e. cognitive ability) is indicated by the accuracy of task responses (e.g. the number of general knowledge questions answered correctly, reflecting semantic knowledge). 'Meta-level' performance (i.e. metacognitive monitoring ability) is indicated by the extent of correspondence between a participant's *judgement* of performance and their *actual* performance (usually established by gamma correlations; Goodman and Kruskal, 1954). One metacognitive judgement widely investigated in the neurotypical population is a judgement of confidence (JOC). In a JOC task, participants often answer general knowledge questions and, immediately after answering a question, are asked to make a judgement about the likelihood that their answer is correct. Higher confidence in correct answers than incorrect answers (a positive gamma) indicates that a person is accurately monitoring their own state of knowledge. Five studies have explored the accuracy of JOC in *children* with ASD, three of which have observed diminished accuracy relative to comparison participants (Grainger et al., 2016; McMahon et al., 2016; Wilkinson et al., 2010; but see Sawyer et al., 2014 and Wojcik et al., 2011).

One effect studied in relation to JOC accuracy (but which has never been explored in ASD) is the 'hypercorrection effect'. When young neurotypical adults (e.g. Butterfield and Metcalfe, 2006) and children (Metcalfe and Finn, 2012) report high confidence in answers that turn out to be incorrect, their memory for the correct answer is subsequently enhanced, suggesting that monitoring of one's confidence mediates learning. In the typical paradigm, participants answer general knowledge questions and, immediately after answering each question, are asked to make a judgement about the likelihood that their answer is correct (just as in a standard JOC task). After each confidence judgement has been made, participants are presented with the correct answer to the question. After all trials have been completed, participants are retested on all the questions that they answered *in*correctly during the initial test phase. During this retest phase, participants are significantly more likely to answer questions correctly if the answer they provided to that question in the initial test phase had been made with high confidence than if it had been made with low confidence. This finding is remarkably robust and has been found reliably in young adults and children from the age of around 6 years (Butler et al., 2011; Butterfield and Mangels, 2003; Butterfield and Metcalfe, 2001, 2006; Eich et al., 2013; Fazio and Marsh, 2009, 2010; Iwaki et al., 2013; Metcalfe et al., 2012; Metcalfe and Finn, 2011, 2012; Metcalfe and Miele, 2014; Sitzman et al., 2014, 2015).

An influential explanation for the hypercorrection effect is that participants experience surprise at being wrong when they did not expect to be wrong and subsequently allocate attentional resources preferentially to encode the correct answer more effectively than answers that they did not expect to be correct during the initial phase (Butterfield and Mangels, 2003; Butterfield and Metcalfe, 2006; Fazio and Marsh, 2009; Metcalfe and Finn, 2011). In other words, corrective feedback following a high-confidence error produces a 'metacognitive mismatch' (e.g. Metcalfe et al., 2012) between one's belief about reality and reality itself. Becoming aware of this mismatch (i.e. of one's own false beliefs) is a phenomenally salient experience that results in deeper processing of the stimuli about which one held a false belief, according to this explanation. The possibility that the hypercorrection effect reflects metacognitive processes is supported indirectly by the finding (by Metcalfe et al., 2012) that corrective feedback to high-confidence, but not low-confidence, errors selectively elicits activation of the right temporo-parietal junction (TPJ, as well as dorsolateral prefrontal cortex). Crucially, the TPJ is a brain region widely implicated in mental state reasoning and metarepresentation of others' thoughts (see Schurz et al., 2014).

However, it may be that the hypercorrection effect reflects a cognitive, rather than (or in addition to) a metacognitive, process. The fact that a metacognitive mismatch (between one's belief about reality and reality itself) enhances learning does not show that the cause of the hypercorrection effect is itself metacognitive, and the above evidence for the involvement of metacognition in producing this effect is only indirect. Instead, the hypercorrection effect may reflect, at least partly, the detection of a mismatch between the generated response and the subsequent feedback. At the time of receiving feedback, participants will of course be aware of what answer they

just gave and whether they reported high or low confidence in this answer. A high-confidence error may thus still elicit a salient mismatch that triggers attentional orienting and enhanced encoding of the correct answer, but the locus of the mismatch may be between the corrective feedback and a representation of the preceding response in short-term memory, rather than a mismatch between corrective feedback and a metacognitive representation of one's previously held false belief (for a discussion of this issue, see Williams and Happé, 2009).

Whether or not the hypercorrection effect reflects metacognitive or cognitive processes, it is important to investigate it in relation to ASD, given the potential implications it has for educational practice. To date, no study has investigated the hypercorrection effect in ASD. If children with ASD show a normal hypercorrection effect, then it might prove a useful learning strategy for teachers to encourage children to make confidence judgements before they receive feedback. However, while neurotypical children (from middle childhood onwards, at least) show a normal hypercorrection effect, it is not clear that children with ASD will. The reaction of children with ASD to a metacognitive mismatch may be atypical, and so attentional resources may not be rallied to encode the correct answers to high-confidence errors more effectively than correct answers to low-confidence errors. If the hypercorrection effect *does* rely on metarepresentational resources, then we should expect the effect to be reduced in ASD (given well-known metarepresentational difficulties in this disorder). However, if the hypercorrection effect merely represents the detection of mismatch at the cognitive level, then the effect may be undiminished in ASD. In Experiment 2, we address the issue (among others) of whether children with ASD display a typical hypercorrection effect. Before this, however, Experiment 1 addresses the issues concerning the nature of the hypercorrection effect and whether JOC accuracy (metacognitive monitoring ability) relates to this effect or to mindreading or ASD-like traits. The results of this first experiment inform predictions for Experiment 2.

In Experiment 1, we investigated JOC accuracy and the hypercorrection effect using the standard task described above. In total, 83 neurotypical adults completed this task, as well as widely used measures of mindreading (the Reading the Mind in the Eyes (RMIE) task; Baron-Cohen et al., 2001a) and ASD-like traits (the autism spectrum quotient (AQ); Baron-Cohen et al., 2001b). Experiment 1 had several related aims.

First, we aimed to assess the extent to which objectively measured metacognitive monitoring ability (JOC accuracy) is associated with mindreading ability. This is the subject of a theoretical dispute between those who argue that metacognitive monitoring depends on the same metarepresentational resources as mindreading (e.g. Carruthers, 2011) and those who claim that metacognitive monitoring and mindreading rely on entirely different processes (e.g. Nichols and Stich, 2003). Mindreading clearly relies on metarepresentation, so if either JOC accuracy or the hypercorrection effect is metarepresentational, then it should be associated with performance on a measure of mindreading ability (in this case the RMIE task).

Second, we aimed to establish the extent to which metacognitive monitoring ability is associated with the number of self-reported ASD-like traits. It is well established that mindreading ability is (negatively) associated with ASD-like traits in the general population (e.g. Baron-Cohen et al., 2001b), but no study to our knowledge has assessed the possible relation between metacognitive monitoring and ASD-like traits, despite suggestions that diminished metacognitive monitoring might contribute to some of the core features of ASD itself (Williams, 2010).

Experiment 1: method

Participants

In total, 83 undergraduate students (72 females) from the University of Kent, United Kingdom, took part in the experiment. The average age of participants was 20.49 years (standard deviation (SD) = 4.96; range = 18–44 years). No participant had a history of psychiatric disorder, including ASD, according to self-report. All participants gave informed consent and received course credit in partial fulfilment of their degree, for taking part in the study. The study (comprising Experiments 1 and 2) was ethically approved by Kent School of Psychology Research Ethics Committee.

Materials and procedures

Hypercorrection task. Stimuli for the hypercorrection task were 150 general knowledge questions (e.g. 'What is the largest planet in the solar system?') taken from Nelson and Narens' set of published questions (Nelson and Narens, 1980) and general knowledge trivia websites. Correct answers to these questions were always a single word (e.g. 'Jupiter'). Questions were presented in a fixed random order for each participant.

The procedure for the hypercorrection task consists of an initial test phase and an unexpected retest phase. The procedure used was based closely on the typical method used to assess the hypercorrection effect in the literature (see, for example, Butterfield and Metcalfe, 2006; Metcalfe et al., 2012). During the initial test, participants were individually presented with each of the general knowledge questions on a laptop screen and were given an unlimited amount of time to provide an answer. Participants were told that if they did not know the answer to the questions they should take a guess. For each question, after participants had provided an answer, a confidence scale appeared below the question and their given

answer. This consisted of a sliding scale, ranging from 'Not Confident' (0) to 'Confident' (100). Participants were instructed to drag the marker left or right with a mouse to make a confidence rating.

Immediately after providing a confidence rating, participants were provided with feedback and were shown the correct answer to the question on the screen for 3 s. It was explained to participants that if the correct answer matched the answer participants had given, feedback would appear in green in the centre of the screen. In contrast, if the correct answer did not match the participant's response, the correct answer appeared in red in the centre of the screen. A letter-matching algorithm was used to score participants' responses as either correct or incorrect automatically. This algorithm calculated the Jaro–Winkler distance between participant's response and the correct answer and did not take into account differences in capital letter use (e.g. 'Jupiter' was considered the same as 'jupiter' or 'JUPITER'). Answers that had a Jaro–Winkler distance ⩾0.80 were considered correct, while answers with a Jaro–Winkler distance ⩽0.79 were considered incorrect. Pilot work indicated that this algorithm classified response correctly on almost all occasions. On the very rare occasion that the letter-matching algorithm misclassified a response as incorrect/correct, the experimenter verbally corrected the feedback. Before completing the initial test phase, participants completed five practice trials.

After participants had completed the initial test phase of the task, they were given a surprise retest. It was explained to participants that they would now be retested on the questions that they had answered incorrectly during the initial test stage. During the retest phase, participants were presented with each of the questions they had answered incorrectly again and were asked to provide an answer. Participants did not have to rate their confidence during the retest phase and were not given feedback on whether their answers were correct.

RMIE task. The RMIE task is a widely used measure of mindreading in clinical and non-clinical populations. Participants were presented with a series of 36 photographs of the eye-region of the face. On each trial, participants were asked to pick one word from a selection of four to indicate what the person in the picture was thinking or feeling. Participants were instructed to read all four words carefully before they made a response. If participants felt more than one of the words was applicable, they were instructed to select the word that they thought was most suitable. Before beginning the task, participants completed a practise trial. Stimuli were presented on screen to participants in a random order, and no time limit was imposed.

AQ. All participants completed the AQ, a self-report questionnaire that assesses ASD/ASD-like features, and is considered a reliable measure of ASD traits in both clinical and subclinical populations. The AQ presents participants' individual statements (e.g. 'I find it difficult to imagine what it would be like to be someone else'), and participants were asked to decide the extent to which they agreed with each statement, responding on a 4-point Likert scale, ranging from 'definitely agree' to 'definitely disagree'.

Scoring

Hypercorrection task. *Object-level* performance (basic general knowledge) ability was calculated as the proportion of questions participants answered correctly in the initial test phase of the task.

Metacognitive monitoring performance (JOC accuracy) was established using gamma correlations, which are a non-parametric measure of the strength of association (Goodman and Kruskal, 1954). Gamma is used in the majority of studies of metacognitive monitoring (and all studies of JOC accuracy within studies of the hypercorrection effect; see Nelson, 1984). A gamma score was obtained for each participant, allowing the association between correctness of each answer *during the initial test phase* and confidence level in the answer given. A gamma score of +1 indicates a perfect correspondence between confidence in the accuracy of answers and actual accuracy of those answers (i.e. a person knows perfectly when they know and do not know). A gamma of 0 indicates no association between confidence and answer accuracy (i.e. random judgements).

The *hypercorrection effect* was measured by calculating a gamma correlation for each participant between confidence assigned to *incorrect* answers *during the initial test* and correctness of answer *during the retest phase* (i.e. whether or not the initial error was corrected at retest). In this case, a gamma of +1 would indicate that initial confidence perfectly predicted retest performance (the higher the confidence in the original error, the more likely it was to be corrected at retest). A gamma of 0 indicates no association between initial confidence and likelihood of error correction (i.e. random corrections). This approach to scoring is employed in the majority of studies of the hypercorrection effect (Butler et al., 2011; Butterfield and Metcalfe, 2001, 2006; Eich et al., 2013; Fazio and Marsh, 2009, 2010; Iwaki et al., 2013; Metcalfe and Finn, 2011, 2012; Metcalfe and Miele, 2014; Metcalfe et al., 2012; Sitzman et al., 2014, 2015).

RMIE task. Scores on the RMIE task range from a possible 0 to 36, with higher scores indicating better performance on the task.

AQ. Scores on the AQ range from 0 to 50, with higher scores indicating more self-reported ASD-like traits. A score ⩾26 is considered as potentially clinically

Table 1. Descriptive statistics for performance on the experimental metacognition/hypercorrection task in Experiment 1.

Variable	Mean (SD)
Initial test phase	
Proportion of correct answers	0.52 (0.15)
Mean confidence rating for all answers	49.22 (15.40)
Judgement of confidence accuracy (gamma)	0.86 (0.06)
Retest phase	
Proportion of correct answers on retest test	0.74 (0.15)
Hypercorrection effect (gamma)	0.25 (0.28)

SD: standard deviation.

significant and expected for a person with a diagnosis of ASD (see Woodbury-Smith et al., 2005).

Statistical analysis

An alpha level of 0.05 was used to determine statistical significance, unless experimental effects were specifically predicted. If effects were predicted, then significance values are reported for one-tailed tests. Where analyses of variance (ANOVAs) were used, we report partial eta squared (η_p^2) values as measures of effect size ($\geqslant 0.01 =$ small effect, $\geqslant 0.06 =$ moderate effect, $\geqslant 0.14 =$ large effect; Cohen, 1969). Where t-tests were used, we report Cohen's d values as measures of effect size ($\geqslant 0.20 =$ small effect, $\geqslant 0.50 =$ moderate effect, $\geqslant 0.80 =$ large effect; Cohen, 1969).

Experiment 1: results

The average AQ total score across participants was 16.08 (SD = 6.20; range = 4–29), which is non-significantly different from the general population average of 16.40 (SD = 6.30) reported by Baron-Cohen et al. (2001b), $t(82) = 0.46$, $p = 0.64$, $d = 0.05$. The average score on the RMIE task was 25.56 (SD = 3.93; range = 4–29), which is non-significantly different from the general population average of 26.20 (SD = 3.60) reported by Baron-Cohen et al. (2001a), $t(82) = 1.45$, $p = 0.14$, $d = 0.19$. These analyses suggest that the current sample is highly representative of the general population with respect to self-reported ASD-like traits and objectively measured mindreading ability.

Descriptive statistics for variables associated with performance on the experimental metacognition/hypercorrection task are shown in Table 1. Participants were highly accurate in their JOCs, producing an average JOC gamma that was significantly above zero, $t(82) = 137.04$, $p < 0.001$. Likewise, participants showed the expected hypercorrection effect, producing an average hypercorrection gamma that was significantly above zero, $t(82) = 0.34$, $p < 0.001$.[1] Moreover, the size of the hypercorrection gamma in this study (0.25) was non-significantly different to the average hypercorrection gamma found across all previous studies of this effect among young neurotypical adults (mean gamma across 24 experiments/independent samples = 0.27; Butler et al., 2011; Butterfield and Metcalfe, 2001, 2006; Eich et al., 2013; Fazio and Marsh, 2009, 2010; Iwaki et al., 2013; Metcalfe and Finn, 2011, 2012; Metcalfe and Miele, 2014; Metcalfe et al., 2012; Sitzman et al., 2014, 2015), $t(82) = 0.59$, $p = 0.55$, $d = 0.09$.

Association analyses

A series of correlation analyses was conducted exploring the relations among performance on the experimental metacognition/hypercorrection task, the RMIE task and the AQ. As predicted, JOC gamma was also *positively* and significantly associated with RMIE, reflecting an association between metacognitive monitoring accuracy and mindreading abilities, $r = 0.25$, $p = 0.01$ (one-tailed). The correlation between JOC gamma and the hypercorrection effect gamma was also positive and significant, $r = 0.32$, $p = 0.003$. However, the hypercorrection effect gamma was negatively and non-significantly associated with the RMIE, $r = -0.16$, $p = 0.88$. A Fisher's Z-test revealed that the JOC gamma × RMIE correlation was significantly larger than the hypercorrection effect gamma × RMIE correlation, $Z = 2.64$, $p = 0.008$.

These results appear to suggest that the variance shared between JOC gamma and hypercorrection effect gamma does not overlap with the variance shared between JOC gamma and RMIE score. To investigate these results further, a series of partial correlations was conducted. The JOC gamma × RMIE score, *controlling for hypercorrection effect gamma*, remained significant and increased slightly in magnitude relative to the bivariate correlation, $r = 0.27$, $p < 0.01$ (one-tailed). Similarly, the JOC gamma × hypercorrection effect gamma, *controlling for RMIE score*, remained significant and increased slightly in magnitude relative to the bivariate correlation, $r = 0.34$, $p < 0.002$. Finally, the hypercorrection effect gamma × RMIE score, *controlling for JOC gamma*, remained negative and non-significant, $r = -0.11$, $p = 0.35$. These analyses confirm that the association between JOC gamma and hypercorrection effect gamma has a different underlying basis to the association between JOC gamma and RMIE score. Finally, a partial JOC gamma × RMIE correlation, controlling for the proportion of questions answered correctly during the initial test phase (i.e. general semantic knowledge) and the hypercorrection effect gamma, remained significant, $r = 0.23$, $p < 0.02$ (one-tailed). This final analysis highlights that the variance shared uniquely between JOC gamma and RMIE score reflected metacognitive/metarepresentational, rather than cognitive/representational, processes.
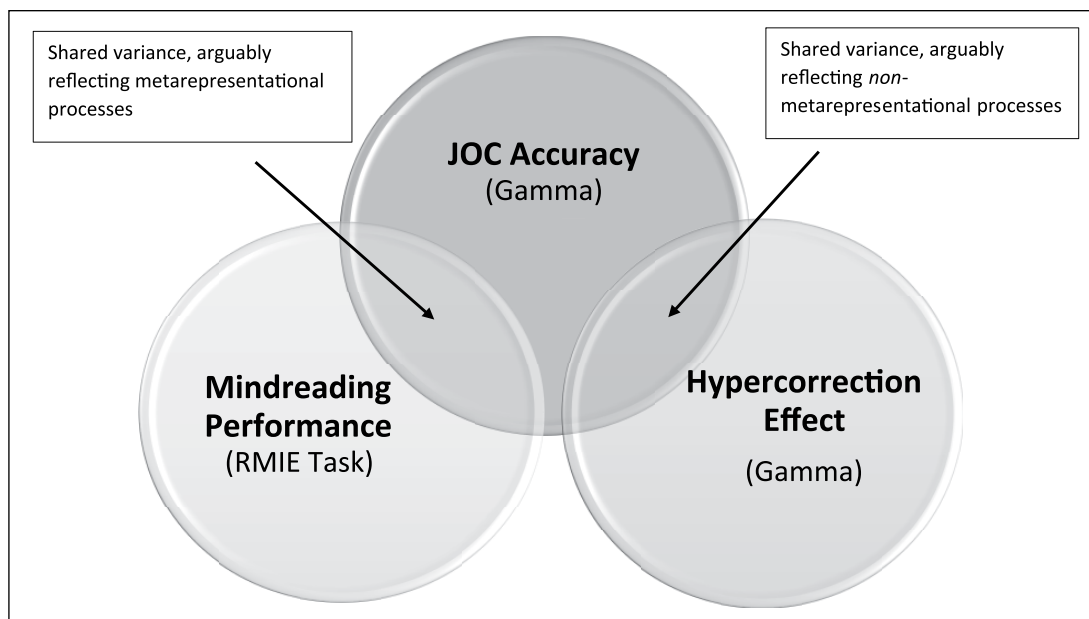
Figure 1. Illustration of the relations among JOC accuracy, RMIE and the hypercorrection effect in Experiment 1.

Next, the relations with AQ were explored. As expected on the basis of previous research, performance of the RMIE task was found to be significantly associated with AQ, $r = -0.35$, $p < 0.001$. However, the hypercorrection effect gamma was non-significantly associated with AQ, $r = 0.08$, $p = 0.45$. Moreover, the JOC gamma × AQ score correlation was non-significant also, $r = -0.10$, $p = 0.36$, suggesting that (unlike mindreading ability) objectively measured metacognitive monitoring ability is not reliably associated with ASD-like traits. To establish the extent to which variance shared between JOC gamma and RMIE is predictive of AQ score, two partial correlations were conducted to establish the extent to which the variance shared between the two is predictive of AQ score. The JOC gamma × AQ score correlation, *controlling for RMIE score*, remained non-significant and reduced in magnitude relative to the bivariate correlation, $r = -0.02$, $p = 0.88$. In contrast, the RMIE × AQ score correlation, *controlling for JOC gamma*, remained highly significant, $r = -0.33$, $p = 0.002$. Thus, what little association there was between JOC accuracy and AQ score was explained almost entirely by a common factor shared with mindreading ability. Furthermore, a Fisher's Z-test confirmed that the partial RMIE × AQ correlation was significantly larger than the JOC gamma × AQ correlation, $Z = 2.02$, $p = 0.04$.

Experiment 1: discussion

The current sample was representative of the population in terms of overall level of performance on widely used measures of mindreading and ASD-like traits. Moreover, the size of the hypercorrection gamma was also representative among the current sample. This is important because

it increases confidence that the observed associations between key dependent variables are also reliable.

The first notable set of findings was that JOC gamma was associated significantly *both* with the hypercorrection effect and with mindreading ability (the RMIE task score). However, the hypercorrection gamma was not, itself, associated with RMIE score. These results suggest that metacognitive monitoring (as indexed by JOC gamma) relies to some extent on metarepresentation (hence the association with a putative metarepresentational task, the RMIE task). However, the variance shared between the hypercorrection effect and JOC accuracy does not reflect a shared reliance on metarepresentation. Rather, it seems that the overlap between JOC accuracy and the hypercorrection effect is due to a non-metarepresentational aspect/component of metacognitive monitoring (see Figure 1 for an illustration). This was confirmed by the partial correlation analyses, in which the relation between JOC gamma and RMIE was even stronger after controlling for the hypercorrection effect gamma, and (correspondingly) the relation between JOC gamma and hypercorrection effect gamma was even stronger after controlling for RMIE score.

The second notable set of findings was that, while RMIE task performance was significantly associated with AQ score (poorer mindreading = more ASD-like traits), neither JOC accuracy nor the hypercorrection effect was predictive of AQ score. The small amount of variance in AQ score explained by JOC accuracy was purely the result of shared variance with mindreading ability. This suggests that, although metacognitive monitoring and mindreading are themselves related, ASD-like traits are related to difficulties in representing *others*' mental states, rather than one's own. However, it is not always possible/accurate to

extrapolate data from studies of ASD-like traits in the general population and apply it to individuals with a full diagnosis of ASD. Although there is arguably continuity between ASD-like behavioural traits in the population and ASD features in diagnosed cases (e.g. Frazier et al., 2014), there may also be qualitative differences in the cognitive mechanisms that underpin those traits in each population (e.g. Peterson et al., 2005). Thus, in our Experiment 2, we addressed these issues in a sample of children with ASD and neurotypical comparison children. Participants completed an age-appropriate version of the experimental hypercorrection task used in Experiment 1, which required them to answer general knowledge questions, provide confidence judgements about each answer, receive feedback on the accuracy of their answer and then be retested on every initially incorrect answer.

Given that JOC accuracy is associated significantly with mindreading ability (the latter being known to be impaired in ASD), it was predicted that participants with ASD would show a significant diminution in JOC accuracy (in line with some previous findings; Grainger et al., 2016; McMahon et al., 2016; Wilkinson et al., 2010), as a result of diminished metarepresentational ability. Making predictions about between-group differences in the hypercorrection effect is less straightforward. The significant association between the JOC gamma and the hypercorrection gamma might lead to the prediction that *both* will be impaired in ASD, given that the two share variance. However, this shared variance is not underpinned by metarepresentational ability, it seems. Rather, the hypercorrection effect does not seem to rely on metarepresentation, unlike JOCs. As such, the prediction that both JOC accuracy and the hypercorrection effect will be diminished in ASD relies on the assumption that the *non*-metarepresentational component of both is impaired in ASD. This may be the case. However, our central prediction (based on background theory; Carruthers, 2009) in this study was that only metarepresentational aspects of the experimental task would be diminished in ASD. For this reason, we predicted that the hypercorrection effect would be undiminished in ASD. Nonetheless, given the uncertainties in making predictions, all *p* values regarding the hypercorrection effect were reported two-tailed to reflect a non-directional hypothesis.

Experiment 2: method

Participants

Baseline characteristics for ASD and comparison participants are displayed in Table 2. All participants with ASD had verified diagnoses made according to *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; DSM-IV-TR) criteria. Verbal intelligence quotient (IQ) and performance IQ were estimated using the Wechsler Abbreviated Scale for Intelligence-II (Wechsler, 2011). The parents of

Table 2. Baseline characteristics of ASD and comparison participants in Experiment 2.

	ASD (*n* = 11; 10 males)	Comparison (*n* = 11; 8 males)
Age	9.86 (1.69)	9.86 (1.00)
VIQ	98.91 (15.34)	104.82 (9.94)
PIQ	104.55 (14.51)	109.91 (14.34)
SRS	78.73 (9.40)	53.44 (18.11)

ASD: autism spectrum disorder; VIQ: verbal intelligence quotient; PIQ: performance intelligence quotient; SRS: Social Responsiveness Scale.

ASD and comparison participants completed the Social Responsiveness Scale (SRS), a valid and reliable measure of ASD features (Constantino and Gruber, 2012). There were no significant between-group differences in age, verbal IQ or performance IQ (all $ps > 0.17$, all $ds < 0.47$). The between-group difference in SRS score was highly significant, $p < 0.001$, $d = 1.75$.

Materials, procedures and scoring

The experimental hypercorrection task had exactly the same structure, procedure and scoring method as that used in Experiment 1, but used different (age-appropriate) stimuli. Three sets of general knowledge questions were used. Each set consisted of 50 questions, taken from Nelson and Narens' published set (Nelson and Narens, 1980) and from general knowledge trivia websites. Each set of questions varied in difficulty and were designed to test the general knowledge of children aged 7–8 years (Set 1), 9–11 years (Set 2) and 12–14 years (Set 3). Given the age range of children who participated in the study, this ensured that the questions in the task would always range in difficulty. This design maximised variation in the confidence ratings assigned to answers (including incorrect answers) and avoided floor or ceiling effects in object-level performance. The decision surrounding which set of questions to give to each child was based on chronological age. However, on the very few occasions when a child's verbal abilities were well above or below average, the question set was determined by mental age to minimise the risk of floor or ceiling effects. Answers to questions in all sets were a single word. Questions were presented in a fixed random order for each participant.

Experiment 2: results

Descriptive statistics for variables associated with performance on the experimental metacognition/hypercorrection task are shown in Table 3.

Object-level performance

Object-level performance was assessed using a mixed 2 (Group: ASD/comparison) × 2 (Test period: initial test/

Table 3. Descriptive statistics for performance on the experimental metacognition/hypercorrection task among ASD and comparison participants in Experiment 2.

Variable	ASD	Comparison
Initial test phase		
Proportion of correct answers	0.54 (0.14)	0.65 (0.11)
Mean confidence rating for all answers	66.94 (22.71)	70.84 (10.51)
Accuracy of confidence judgements (gamma)	0.74 (0.17)	0.86 (0.12)
Retest phase		
Proportion of correct answers on retest test	0.62 (0.21)	0.76 (0.10)
Hypercorrection effect (gamma)	0.47 (0.47)	0.35 (0.50)

ASD: autism spectrum disorder.

retest) ANOVA on the proportion of questions answered correctly. This revealed a significant main effect of Test period, reflecting significantly superior performance at retest than at initial test, $F(1, 20) = 10.91$, $p = 0.004$, $\eta_p^2 = 0.35$. The main effect of Group was also significant, reflecting superior semantic knowledge among comparison participants than ASD participants overall, $F(1, 20) = 5.04$, $p = 0.04$, $\eta_p^2 = 0.20$. However, there was no Group × Test period interaction, showing that participants with and without ASD showed the same pattern of performance across test periods, $F(1, 20) = 0.21$, $p = 0.65$, $\eta_p^2 = 0.01$.

JOCs and the hypercorrection effect

The between-group difference in average confidence ratings given to answers during the initial test phase was non-significant, $t(14.10) = 0.52$, $p = 0.61$, $d = 0.22$. JOC *accuracy* (gamma) was significantly above zero in each group, $ts \geqslant 14.33$, $ps \leqslant 0.001$ (one-tailed). However, as predicted, JOC gamma was significantly lower among participants with ASD than comparison participants, $t(20) = 2.00$, $p = 0.03$ (one-tailed), $d = 0.86$. This between-group difference is very similar in magnitude to the average between-group difference in JOC accuracy across the five previous studies among children with ASD (mean Cohen's $d = 0.81$; Grainger et al., 2016; McMahon et al., 2016; Sawyer et al., 2014 (Experiments 1 and 2); Wilkinson et al., 2010 (Experiment 1); Wojcik et al., 2011).[2] Finally, the hypercorrection effect gamma was significantly above zero in each group, $ts \geqslant 2.28$, $ps \leqslant 0.02$.[3] Despite showing a substantial diminution of JOC accuracy, participants with ASD nonetheless showed a slightly (but non-significantly) *larger* hypercorrection effect than did comparison participants, $t(20) = 0.57$, $p = 0.58$, $d = 0.25$.

To put these key results in context, participants with ASD performed 1.02 SDs below the comparison group mean for JOC accuracy, but 0.24 SDs above the comparison group mean for the hypercorrection effect, on average. This difference is highly significant, indicating that, relative to the performance of comparison participants, the performance of participants with ASD was significantly

poorer with respect to JOC accuracy than with respect to the hypercorrection effect, $t(10) = 3.01$, $p = 0.01$.

Finally, a series of exploratory correlation analyses was conducted to explore the extent to which performance on the experimental metacognition/hypercorrection task was associated with ASD features (SRS total score). Among ASD participants, the JOC gamma × SRS correlation was moderate-to-large, but non-significant, $r = -0.43$, $p = 0.18$. The hypercorrection effect gamma × SRS correlation was also moderate, but non-significant, $r = -0.31$, $p = 0.35$. Among comparison participants, correlations were small and non-significant (JOC gamma × SRS: $r = -0.10$, $p = 0.80$; hypercorrection effect gamma × SRS: $r = 0.13$, $p = 0.74$).

Experiment 2: discussion

As expected, JOC accuracy was found to be diminished among this sample of children with ASD. The between-group difference in the size of JOC gamma was significant and large ($d = 0.86$). In contrast, the hypercorrection effect was slightly (but non-significantly) larger among ASD than comparison participants. Results indicated that the between-group difference in JOC accuracy was significantly larger than the between-group difference in the hypercorrection effect. One important issue that should be considered when interpreting the results is the extent to which the results are reliable, given the small sample size. This is particularly true when interpreting null effects, given the potential for type II errors when sample sizes are small. An increasingly used approach to interpreting null results is to calculate a Bayes factor associated with the critical contrast of interest. Bayes factors overcome the limitations of interpreting p values when power is relatively low by providing an estimation of the relative strength of a finding for one theory over another theory, which allows a more graded interpretation of the data (e.g. Rouder et al., 2009). Thus, they provide an estimate of the degree to which findings are supportive of the null hypothesis over the alternative hypotheses (Dienes, 2014). According to Jeffreys' (1961) criteria for interpreting Bayes factors, values 3 provide evidence for the alternative hypothesis, whereas values 1 (or ⩽0.33, according to Dienes, 2014) provide evidence for the null hypothesis. Bayes factors

between 1 and 3 provide inconsistent evidence for either hypothesis. We calculated a Bayes factor (using Dienes, 2008) for the critical between-group difference in the hypercorrection gamma (mean difference between groups = −0.12; standard error (SE) of the difference = 0.20), assuming the difference would be associated with an effect size of 0.60.[4] This calculation produced a Bayes factor of 0.21, which indicates moderate evidence for the null hypothesis.[5]

General discussion

The aims of this study were to investigate metacognitive monitoring ability and its relation to ASD features/ASD-like traits and mindreading ability. We also aimed to investigate the hypercorrection effect. In Experiment 1, we found that JOC accuracy – an established indicator of metacognitive monitoring ability – was associated significantly with performance on the RMIE task, which is a widely employed measure of mindreading ability. To our knowledge, this is the first study of the relation between metacognitive judgement accuracy and mindreading among neurotypical individuals, and the results are important from a theoretical perspective. Although previous studies have attempted to address the potential link between metacognition and mindreading among individuals with a formal diagnosis of ASD, those studies have lacked the statistical power required to detect the predicted effect (see Grainger et al., 2014). This is common in ASD research, in which samples tend to be relatively small, given the difficulties recruiting and testing participants with this disorder. By exploring this issue in relation to ASD-like traits and mindreading ability measured as continuous variables among a large sample of neurotypical adults, we were able to overcome this problem. The significant association between JOC accuracy and performance on the RMIE task in Experiment 1 was not substantial ($r = 0.25$; $r = 0.27$ after controlling for hypercorrection effect gamma), but this was expected. Theoretical claims that metacognitive monitoring depends on the same neurocognitive mechanism as mindreading (e.g. Carruthers, 2011; Frith and Happé, 1999; Williams, 2010) do not imply that the two abilities are synonymous or that they are entirely overlapping. Rather, the suggestion is merely that metacognitive monitoring relies to some extent on the same inferential capacity/metarepresentational resources as does mindreading, rather than being an entirely non-inferential, non-metarepresentational process. There are other non-shared resources required for successful mindreading and metacognitive monitoring, respectively; executive functioning, for example, is probably implicated to a greater degree in mindreading performance than metacognitive monitoring performance (Roebers et al., 2012). As such, the finding that the mindreading and metacognitive abilities are related, but not perfectly/substantially so, is in keeping with predictions.

As expected, we also found that mindreading ability was associated significantly with ASD-like traits; the lower an individual's mindreading ability, the higher their ASD-like traits tended to be. However, JOC accuracy was not significantly associated with ASD-like traits in Experiment 1. This suggests that it is primarily the use of metarepresentational resources to represent others' mental states, rather than one's own mental states, which contributes to ASD-like traits in the general population.

Overall, the results from Experiment 1 were complemented by those from Experiment 2. In Experiment 2, we found diminished JOC accuracy, but an undiminished hypercorrection effect, among children with a formal diagnosis of ASD. The finding of diminished JOC accuracy in ASD adds to recent findings that children with ASD have difficulty making this kind of metacognitive judgement (Grainger et al., 2016; McMahon et al., 2016; Wilkinson et al., 2010; but see Sawyer et al., 2014) and supports the idea that individuals with this disorder have diminished metacognitive monitoring ability alongside diminished mindreading ability (see also Cooper et al., 2016).

In this study, despite showing diminished JOC accuracy, children with ASD nonetheless showed an undiminished hypercorrection effect (e.g. Butterfield and Metcalfe, 2006). Likewise, in Experiment 1, whereas JOC accuracy correlated with mindreading ability, the hypercorrection effect did not. Although a prominent theoretical framework suggests that the hypercorrection effect is caused by metacognitive mismatch detection, our findings suggest that this effect may not be strongly dependent on metarepresentational ability, but rather may be mediated by more basic cognitive processes. Of course, it is important to stress that the results concerning the hypercorrection effect in Experiment 2 need to be considered with some caution, given the small sample involved, and await replication before strong conclusions can be drawn. However, the data itself provided moderate evidence that this effect is undiminished in ASD. This is the first study of the hypercorrection effect in ASD and suggests that children with this disorder would benefit from making explicit confidence judgements about the state of their own knowledge when learning new information. Corrective feedback to high-confidence errors selectively enhances learning and so could help children with ASD to learn information more effectively, even if the JOCs are not themselves as accurate as those made by neurotypical children.

Interestingly, the pattern of performance observed in children with ASD in Experiment 2 is the opposite of that observed among older neurotypical adults. Eich et al. (2013) found that, relative to young adults, older adults showed a significantly diminished hypercorrection effect despite undiminished JOC accuracy. Although the findings from Experiment 2 with regard to the hypercorrection effect are preliminary, they suggest a different pattern of strengths and difficulties in this domain among people

with ASD compared to older neurotypical adults. This is particularly interesting, given recent comparisons between cognitive/memory impairments in ASD and cognitive decline in older age (e.g. Bowler, 2007; Ring et al., 2016), and suggests that this comparison may not be appropriate in all respects.

In sum, this study provides further evidence of metacognitive monitoring deficits in ASD, but also of typical hypercorrection of high-confidence errors in children with this disorder. Future research might usefully take a developmental/longitudinal approach to these issues to establish whether metacognitive monitoring deficits resolve over time in ASD. Moreover, it would be beneficial to know whether individuals with ASD base their same sources of information (or 'mnemonic cues') as neurotypical individuals do. The investigation of metacognition in ASD is at a relatively early stage, so many questions remain unanswered. This research helps narrow down the questions that need to be asked by confirming that metacognitive monitoring deficits are present in children, as well as by clarifying the relation between monitoring ability and ASD traits, and the nature of the hypercorrection effect.

Notes

1. No differences between males and females were apparent in terms of either judgement of confidence (JOC) gamma, $t(81) = 0.76$, $p = 0.45$, $d = 0.25$, or the hypercorrection effect gamma, $t(81) = 0.32$, $p = 0.75$, $d = 0.11$. Moreover, there were no between-sex differences in any of the bivariate correlations reported in Experiment 1 (all $Zs \leqslant 0.93$, all $ps \geqslant 0.35$).
2. Wilkinson et al. (2010) did not use gamma to establish JOC accuracy. The effect size is based on between-group differences in the proportion of items correctly recognised (old/new recognition) that participants were certain they had got correct (see Table 2 of Wilkinson et al., 2010: 1375). Wojcik et al. (2011) did not use gamma to establish JOC accuracy. The effect size is based on the average difference in JOC accuracy (as established using 'non-directional discrepancy method') across the three conditions of Wojcik et al.'s experimental task.
3. In the only other study of the hypercorrection effect in children, Metcalfe and Finn (2012) reported an average

hypercorrection gamma of 0.21 across three experiments involving children aged approximately 8–12 years. The hypercorrection effect gammas in this study were larger than this, but non-significantly so, among participants with autism spectrum disorder (ASD), $t(10) = 1.80$, $p = 0.10$, and comparison participants, $t(10) = 0.90$, $p = 0.39$.
4. This effect size estimate was based on the difference in the size of the hypercorrection effect between younger and older adults in the study by Eich et al. (2013). Given that the memory/cognitive profile of older adults has been compared to that characteristically observed in ASD (see section 'General discussion'; for example, Bowler, 2007), we considered that any between-group difference in the size of the hypercorrection effect in this study might be of a similar magnitude to that observed by Eich et al.
5. An alternative to basing an effect size estimate on existing literature is to specify the 'minimally interesting value' that would support a particular hypothesis (e.g. Dienes, 2014). That is, one can ask what size a between-group difference in task performance would have to be for it to be meaningful/clinically relevant. In our view, any between-group difference in task performance should be associated with an effect size that is at least moderate (i.e. 0.50). Any difference that is only small in magnitude is not clearly clinically relevant, in our view. Therefore, we recalculated the Bayes factor for the predicted between-group difference in the size of the hypercorrection effect, but assumed that the difference would be associated with an effect size of 0.50, rather than 0.60. Under this conservative assumption, the Bayes factor was 0.25, which still represents moderate evidence in favour of the null hypothesis.

References

American Psychiatric Association (APA) (2013) *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: APA.

Baron-Cohen S, Wheelwright S, Hill J, et al. (2001a) 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): 241–252.

Baron-Cohen S, Wheelwright S, Skinner R, et al. (2001b) 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.

Bowler DM (2007) *Autism Spectrum Disorder: Psychological Theory and Research*. Chichester: John Wiley & Sons.

Brosnan M, Johnson H, Grawemeyer B, et al. (2016) Deficits in metacognitive monitoring in mathematics assessments in learners with autism spectrum disorder. *Autism: International Journal of Research and Practice* 20(4): 463–472.

Brunsdon VEA and Happé F (2014) Exploring the 'fractionation' of autism at the cognitive level. *Autism* 18(1): 17–30.

Butler AC, Fazio LK and Marsh EJ (2011) The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review* 18(6): 1238–1244.

Butterfield B and Mangels JA (2003) Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research* 17: 793–817.

Butterfield B and Metcalfe J (2001) Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27: 1491–1494.

Butterfield B and Metcalfe J (2006) The correction of errors committed with high confidence. *Metacognition and Learning* 1: 1556–1623.

Carruthers P (2009) How we know our own minds: the relationship between mindreading and metacognition. *Behavioral and Brain Sciences* 32(2): 121–138.

Carruthers P (2011) *The Opacity of Mind*. Oxford: Oxford University Press.

Cohen J (1969) *Statistical Power Analysis for the Behavioural Sciences*. San Diego, CA: Academic Press.

Constantino JN and Gruber CP (2012) *Social Responsiveness Scale, Second Edition (SRS-2)*. Torrance, CA: Western Psychological Services.

Cooper R, Plaisted-Grant K, Baron-Cohen S, et al. (2016) Reality monitoring and metamemory in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders* 46: 2186–2198.

Dienes Z (2008) *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Palgrave Macmillan.

Dienes Z (2014) Using Bayes to get the most out of non-significant results. *Frontiers in Psychology* 5: 781.

Eich TS, Stern Y and Metcalfe J (2013) The hypercorrection effect in younger and older adults. *Neuropsychology, Development, and Cognition. Section B: Aging, Neuropsychology and Cognition* 20(5): 511–521.

Estes A, Rivera V, Bryan M, et al. (2011) Discrepancies between academic achievement and intellectual ability in higher-functioning school-aged children with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 41(8): 1044–1052.

Fazio LK and Marsh EJ (2009) Surprising feedback improves later memory. *Psychonomic Bulletin & Review* 16: 88–92.

Fazio LK and Marsh EJ (2010) Correcting false memories. *Psychological Science* 21: 801–803.

Frazier TW, Ratliff KR, Gruber C, et al. (2014) Confirmatory factor analytic structure and measurement invariance of quantitative autistic traits measured by the Social Responsiveness Scale-2. *Autism* 18(1): 31–44.

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

Goldman AL (2006) *Simulating Minds*. Oxford: Oxford University Press.

Goodman LA and Kruskal WH (1954) Measures of association for cross classifications. *Journal of the American Statistical Association* 49(268): 732–764.

Grainger C, Williams DM and Lind SE (2014) Metacognition, metamemory, and mindreading in high-functioning adults with autism spectrum disorder. *Journal of Abnormal Psychology* 123: 650–659.

Grainger C, Williams DM and Lind SE (2016) Metacognitive monitoring and control processes in children with autism spectrum disorder: diminished judgement of confidence accuracy. *Consciousness and Cognition* 42: 65–74.

Griffin C, Lombardo M and Auyeung B (2015) Alexithymia in children with and without autism spectrum disorders. *Autism Research* 9: 773–780.

Hartwig MK and Dunlosky J (2012) Study strategies of college students: are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review* 19(1): 126–134.

Iwaki N, Matsushima H and Kodaira K (2013) Hypercorrection of high confidence errors in lexical representations. *Perceptual and Motor Skills* 117: 219–235.

Jeffreys H (1961) *Theory of probability* (Third Edition). Oxford, UK: Oxford University Press.

Jones CR, Happé F, Golden H, et al. (2009) Reading and arithmetic in adolescents with autism spectrum disorders: peaks and dips in attainment. *Neuropsychology* 23(6): 718–728.

McMahon C, Henderson H, Newell L, et al. (2016) Metacognitive awareness of facial affect in higher-functioning children and adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 46(3): 882–898.

Metcalfe J and Finn B (2011) People's correction of high confidence errors: did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 437–448.

Metcalfe J and Finn B (2012) Hypercorrection of high confidence errors in children. *Learning and Instruction* 22: 253–261.

Metcalfe J, Butterfield B, Habeck C, et al. (2012) Neural correlates of people's hypercorrection of their false beliefs. *Journal of Cognitive Neuroscience* 24(7): 1571–1583.

Metcalfe J and Miele DB (2014) Hypercorrection of high confidence errors: prior testing both enhances delayed performance and blocks the return of the errors. *Journal of Applied Research in Memory and Cognition* 3: 189–197.

Nelson TO (1984) A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin* 95(1): 109–133.

Nelson TO and Narens L (1980) Norms of 300 general-information questions: accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior* 19: 338–368.

Nichols S and Stich SP (2003) *Mindreading: An Integrated Account of Pretence, Self-awareness, and Understanding Other Minds*. Oxford: Clarendon Press/Oxford University Press.

Peterson C, Wellman H and Liu D (2005) Steps in theory-of-mind development for children with deafness or autism. *Child Development* 76(2): 502–517.

Ring M, Gaigg SB and Bowler DM (2016) Relational memory processes in adults with autism spectrum disorder. *Autism Research* 9(1): 97–106.

Roebers CM, Cimeli P, Roethlisberger M, et al. (2012) Executive functioning, metacognition, and self-perceived competence in elementary school children: an explorative study on their interrelations and their role for school achievement. *Metacognition and Learning* 7: 151–173.

Rouder JN, Speckman PL, Sun D, et al. (2009) Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin Review* 16: 225–237.

Sawyer AC, Williamson P and Young R (2014) Metacognitive processes in emotion recognition: are they different in adults with Asperger's disorder? *Journal of Autism and Developmental Disorders* 44(6): 1373–1382.

Schurz M, Radua J, Aichhorn M, et al. (2014) Fractionating theory-of-mind: a meta-analysis of functional brain

imaging studies. *Neuroscience & Biobehavioral Reviews* 42: 9–34.

Sitzman DM, Rhodes MG and Tauber SK (2014) Prior knowledge is more predictive of error correction than subjective confidence. *Memory & Cognition* 42: 84–96.

Sitzman DM, Rhodes MG, Tauber S, et al. (2015) The role of prior knowledge in error correction for younger and older adults. *Neuropsychology, Development, and Cognition. Section B: Aging, Neuropsychology and Cognition*. Epub ahead of print 3 December. DOI: 10.1080/13825585.2014.993302.

Veenman MV, Kok R and Blöte AW (2005) The relation between intellectual and metacognitive skills in early adolescence. *Instructional Science* 33(3): 193–211.

Wechsler D (2011) *Wechsler Abbreviated Scale of Intelligence*. 2nd ed. New York: The Psychological Corporation.

Wilkinson DA, Best CA, Minshew NJ, et al. (2010) Memory awareness for faces in individuals with autism. *Journal of Autism and Developmental Disorders* 40(11): 1371–1377.

Williams DM (2010) Theory of own mind in autism evidence of a specific deficit in self-awareness? *Autism* 14(5): 474–494.

Williams D and Happé F (2009) 'What did I say?' versus 'What did I think?': attributing false beliefs to self amongst children with and without autism. *Journal of Autism and Developmental Disorders* 39: 865–873.

Woodbury-Smith M, Robinson J and Baron-Cohen S (2005) Screening adults for Asperger Syndrome using the AQ: diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders* 35(3): 331–335.

Wojcik DZ, Allen RJ, Brown C, et al. (2011) Memory for actions in autism spectrum disorder. *Memory* 19(6): 549–558.