Linking Mindreading and Metacognition:

The Case of Autism Spectrum Disorder

Katie Louise Carpenter

School of Psychology

University of Kent

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Division of Human and Social Sciences at the University of Kent, Canterbury. October 2021.

Declarations, I declare that this thesis is my own work carried out under the normal terms of supervision.

Publications

Chapter three of this thesis has been published:

Carpenter, K. L., Williams, D. M., & Nicholson, T. (2019). Putting your money where your mouth is: examining metacognition in ASD using post-decision wagering. *Journal of autism and developmental disorders*, *49*(10), 4268-4279.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would also like to thank Professor David Williams for his apparently unwavering belief that I could complete a PhD, without him I would never have started on this journey, and I certainly would never have finished it. I have learnt an enormous amount, both personally and professionally, through the supervision and guidance he has provided. I would also like to thank the Economic and Social Research Council for funding of this project, and to all those at the South East Network for Social Sciences for their support and hard work in creating a supportive academic community, especially Paul Newman and Felicity Szesnat who have always been there to address any questions or concerns when needed.

I would also like to thank all of the participants that gave up their time to take part in each of my studies, they were an absolute pleasure to work with and without them I would have no data to analyse and interpret. Thank you to Frank Gasking too for all the technical support he provided and for introducing me to the joys of coding, I certainly wouldn't have learnt everything there is to know about Psychopy if it hadn't been for his patience.

My sincerest thanks goes to my husband, Oliver Carpenter, for showing me it can be done, and for always believing in me and supporting me in so many ways. I appreciate you and what you do every day. A very special thank you goes to my parents, Lesley and Eric Shoubridge, for the endless hours of unpaid childcare and unconditional support, even if you haven't a clue what on earth it is that I have been doing these last four years. Finally, thank you to my little girl, Sophie Carpenter, for being 'my reason why' and reminding me that there is more to life than a PhD. You joined me in the first year and have been the one that gets me up every day (literally) to get on with it, without you I may well have got it done quicker but I certainly wouldn't have had so much joy and light in my life. This thesis is dedicated to you, and I hope that you can find the belief and courage to fulfil all your dreams, no matter how hard they are. Now let’s go and have some mummy and Sophie time!

Abstract

The aim of this thesis is to establish if the ability to attribute mental states to ourselves (metacognition) and others (mindreading) rely on the same underlying processes. To address this question, the thesis presents multiple studies using novel methods to examine the relationship between these two abilities. It also extends the findings beyond a purely relational nature and examines the question of whether there is a dissociation between mindreading and metacognition among autistic people, a group known to have mindreading difficulties. If indeed there is a relationship, and no dissociation is found, then this is in line with the notion that mindreading and metacognition rely on the same underlying processes.

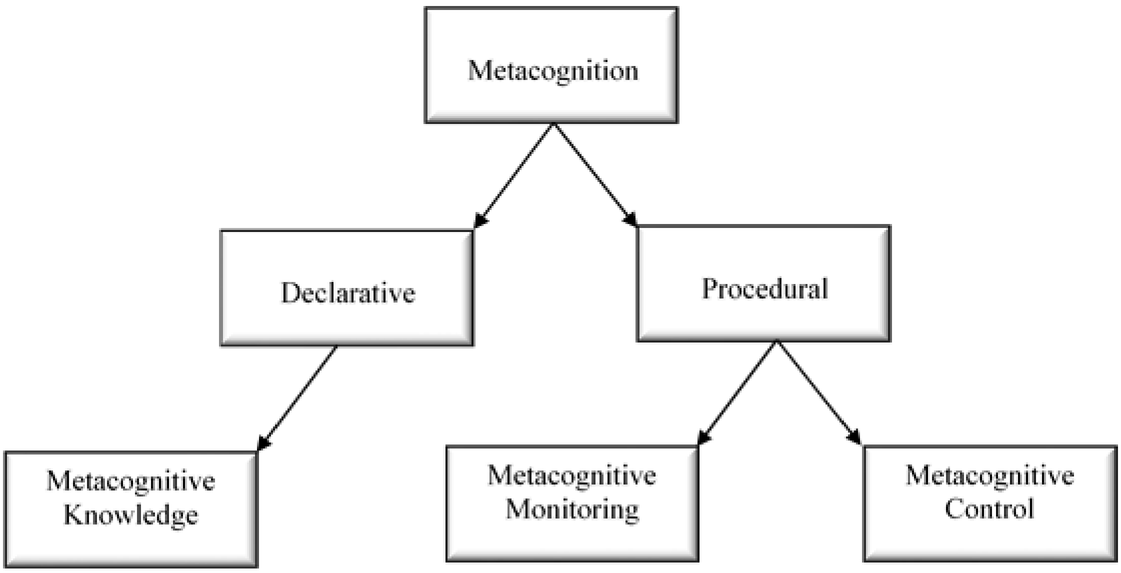
Chapter 1: Introduction

Metacognition

Metacognition is any cognitive process that takes one's own 'cognition' as its subject. It has also been referred to as 'thinking about thinking' or 'cognition about cognition' (Flavell, 1979). Metacognition, enables us to be aware of our own thought processes, and to monitor and control them. It is crucial for everyday functioning and enables us to regulate our own thoughts and behaviours (Nelson & Narens, 1990). It has both declarative and procedural aspects, and has been divided conceptually into three key components: the declarative aspects known as 'metacognitive knowledge', and the procedural aspects known as 'metacognitive monitoring' and 'metacognitive control' (see Figure 1; Flavell, 1979).

Figure 1

Components of Metacognition



Declarative Metacognition

The declarative aspect of metacognition otherwise known as metacognitive knowledge refers to what we know and believe about cognitions and cognitive processes. For example, a person may *know* that they (and/or others) learn better by taking notes as they read, or that a longer list of words is harder to remember than a short list. Metacognitive knowledge has three key elements: person, task and strategy (Flavell, 1979). The person element refers to knowledge about how people process information. For example, knowing how, when and why people remember or forget. The task element includes knowledge about the nature of the cognitive task, and how the characteristic of the task impacts on one's performance. For example, appreciating that a large number of items is harder to recall than a small number of items. The strategy element involves knowledge about different strategies and how they influence task performance, such as knowing that writing notes when reading a textbook is a more effective way to learn the material than just reading it.

Procedural Metacognition

Procedural metacognition includes both metacognitive monitoring and metacognitive control. *Metacognitive monitoring* is how we represent occurrent, ongoing cognitive activity. It is the process of assessing our performance on a particular task. For instance, judging how confident we are that we have learned all we need to pass an exam, or that we know the answer to a question but can't quite bring it to mind. *Metacognitive control*, on the other hand, is our ability to regulate our ongoing cognitive activity, this may include actions such as adjusting how much time we spend revising a particular topic in preparation for an exam. Importantly, metacognitive control is influenced by meta-monitoring. If we judge (meta-monitoring) that we have learnt everything for one exam we can then stop revising for that topic and move on to the next (control).

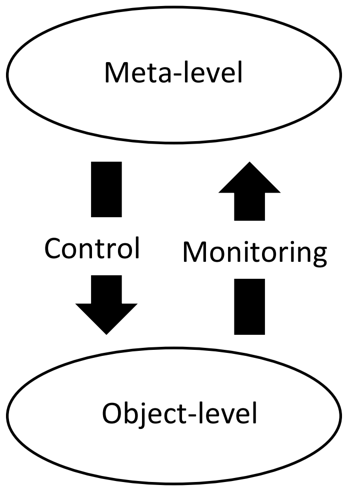
A Conceptual Framework of Metacognition

To explain how metacognitive monitoring and metacognitive control relate to and interact with cognitive processes, Nelson and Narens (1990) developed a highly influential conceptual framework (see Figure 2). They identified two interrelated levels, the 'object-level' and the 'meta-level'. The object-level refers to the ongoing cognitive process, for example, judging the speed of a car coming toward you. The meta-level represents the ongoing object-level activity and is informed by both metacognitive monitoring and metacognitive knowledge. For example, how one represents the object-level judgment, in this case the speed of the car, can be informed by how confident you are that your judgement is correct (metacognitive monitoring), as well as information about the task (metacognitive knowledge), such that it is harder to judge the speed of a car in foggy conditions.

Nelson and Narens highlighted that people only become aware of their cognitive processes when they are represented at the meta-level. Once represented at the meta-level one can then adjust one's cognition and behaviour via metacognitive control, and so the process is both top-down and bottom-up. The meta-level is informed about the object-level via metacognitive monitoring combined with metacognitive knowledge. The object-level is influenced by the meta-level via metacognitive control. Thus, one completes a cognitive task, such as learning a list of words, then information about the task is fed to the meta-level, such as whether one feels they have learnt the list sufficiently (metacognitive monitoring), combined with ones existing metacognitive knowledge, such as how long the list is and what this means in terms of ease of learning. This then enables the individual to adjust their cognition and/or behaviour, and so if the individual has judged that they won't remember everything on the list (metacognitive monitoring) because it is far too long (metacognitive knowledge) they may then decide to continue to study that list or perhaps give up and write the list down instead (metacognitive control).

Figure 2

Nelson & Narens Model of Metacognition (1990)



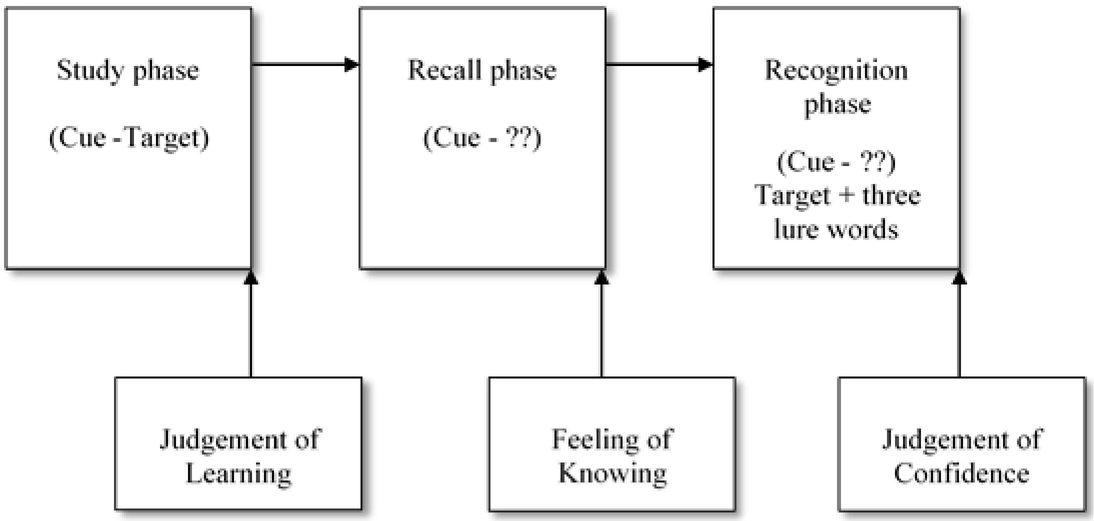
Metacognitive Accuracy

Given that metacognitive control relies so heavily in metacognitive monitoring, it is vital that we *accurately* monitor our own cognitive processes. After all, if we have learnt everything that we need for one exam, but judge that we haven't and so continue to revise that topic we may well run out of time to study for the remaining exams, resulting in suboptimal overall performance. This highlights the importance of metacognition generally but more specifically the importance of accurate metacognitive monitoring.

Metacognitive accuracy can be measured in a variety of ways, with the most commonly used methods being judgements-of-learning, feelings-of-knowing, and judgements-of-confidence. Each of these methods require participants to complete an object-level task, and to make a metacognitive judgment about their performance on that task. A key difference between each measure is the stage at which each judgement is made in relation to the object-level task (see Figure 3). For example, judgements-of-learning and feeling-of-knowing judgements are made *before* completion of the object-level task, whereas judgements-of-confidence are made *after* the object-level task has been completed.

Figure 3

Metacognitive Judgments within a Meta-Memory Experiment



Crucially, the association between the participant's meta-level judgement and their actual performance on the object-level task is used as an indicator of metacognitive accuracy. This is often done by calculating a gamma correlation (Kruskal & Goodman 1954), with scores ranging from − 1 to + 1, where scores of 0 indicate chance level accuracy, and large positive scores indicating good metacognitive accuracy (for a detailed explanation of Kruskal & Goodman gamma correlation, see Chapter two).

Judgement-of-learning, for example, require participants to learn something and then rate how likely it is that they have learnt the target information (meta-level judgement). Following this, participants complete a test of what they have learnt (object-level task). For instance, participants may be asked to learn a list of word-pairs, a cue word and a target word. Participants then rate how likely it is that they will be able to recall each target word during a follow up test. Subsequently participants are presented with the cue word from the previously studied word pairs and asked to recall the target word. The association between the participant's judgement-of-learning for each item and their actual recall for each item is then used to indicate their metacognitive accuracy.

Feeling-of Knowing tasks, on the other hand, require participants to complete a task, make a feeling-of-knowing judgement on any items they get incorrect, and then complete another test for any items they previously got incorrect. For example, participants may be asked to learn a list of word-pairs, a cue and a target word. They are then asked to recall the target word when presented with the cue word. For any items they get incorrect they make a feeling-of-knowing judgement. For example, they may be asked how likely it is that they would recognise the target word if it were presented alongside a number of lure words. After making the feeling-of-knowing judgement participants may be given a recognition test (object-level task). In this case the association between the feeling-of-knowing judgement for each item and the recognition performance for each item indicates the level of metacognitive accuracy.

Judgements-of-confidence also require participants to complete an object-level task, such as recalling a list of previously learnt words or completing a perceptual discrimination task such as deciding which of two images has the most dots. Immediately after each decision participants are asked to rate how confident they are that their answer was correct. As with all previously mentioned tasks the association between the meta-level judgment, in this case the judgement-of-confidence, for each item and the actual performance, such a correct or incorrect discrimination/recall, signifies the accuracy of the participant's metacognition.

In addition to the various verbal measures described above, the measurement of metacognitive accuracy can be assessed using more concrete behavioural measures such as post-decision wagering (Ruffman et al. 2001; Persaud et al. 2007). Like in judgements-of-confidence tasks, participants are required to complete an object-level task followed by their meta-level judgements, only rather than rating their confidence, participants are required to place a wager on their answer. If their object-level answer is correct, then the participant will win what they have wagered, but if they are incorrect, they will lose what has been wagered. So high wagers for correct answers and low wagers for incorrect answers would indicate good metacognitive accuracy. As with the verbal measures, gamma correlations are calculated and used as the metacognitive variable.

Monitoring Metacognition

Given that metacognitive monitoring is central to metacognitive control, the question of *how* one monitors their own cognitive processes has also gained a significant amount of attention. From this debate two main theories have arisen: the direct-access theory (Hart, 1967) and the heuristic-based theory (Koriat, 2007). The direct-access theory (Hart, 1967) claims that people have privileged and direct access to their own mental states. It is argued that when metacognitive judgements are made, one merely detects the presence/strength of the information available to them regarding their mental state. For example, in a memory task people merely detect the strength of the memory in order to decide if they know they will be able to retrieve the information or not.

An alternative explanation is the heuristic-based theory of metacognition (Koriat, 2007). This theory states that people do not have direct access to their own mental states, and that any knowledge or awareness of one's own metal state is inferential in nature. According to this theory, people use cues and heuristics to determine what they do and do not know. This includes the familiarity of a cue as well as how accessible the target information is (Koriat, 1994; Koriat & Levy-Sadot, 2001; Metcalfe, Schwartz, & Joaquim, 1993). For example, in Metcalf et al's study participants were presented with a list of (cue-target) word pairs to learn but the number of times the participants saw the cue word was manipulated. They were then presented with each cue word and asked to rate the likelihood that they would recognise the target word in a future recognition test (feeling-of-knowing judgement). The results showed that the participants made stronger feeling of know judgements on trials where the cue had been presented multiple times during the study phase in comparison to trials where the cue had just been presented just once. Moreover, this relationship was stronger than the relationship between feeling-of-knowing judgement and actual recognition. Metcalfe et al.'s results suggest that people monitor how familiar the cue is, rather than their actual memory of the target information, when making metacognitive judgements.

Further evidence for the heuristic-based theory comes from studies examining the accessibility of the target. Koriat (1994, 2001), for example, found that participants' feeling-of-know judgements became stronger in line with the number of letters recalled from a letter string (e.g. SOACORJCKLC), regardless of the accuracy of the letters recalled. Again, highlighting that the memory itself is not the driving force behind one's metacognitive judgment and that people rely on heuristics, in this case the amount of information recalled regardless of accuracy, to judge their cognitions.

Development of Metacognition

Another key topic of interest in the field of metacognition is the development of metacognitive knowledge, monitoring and control. The majority of research in this area has focused on metamemory (metacognition in relation to memory). Research shows that children as young as four have a partial understanding of verbs such as remember and forget, and by five years of age they display good knowledge and use of such terms (Wellman & Johnson, 1979). Research also shows that five-year-olds have some knowledge of how memory works and strategies that can help improve memory, although this clearly improves with age (Kreutzer, Leonard, Flavell, 1975). For example, Kreutzer et al., found that five-year-olds were aware of the impact that repetition, rehearsal, study time and mnemonics had on one's memory, as well as the impact of the amount of information to be remembered, the delay in recall, and the age of the child (i.e., older children can remember more than younger children).

In terms of metacognitive monitoring and control, children as young as three appear to show some ability to monitor and control their own cognitive processes (Coughlin, Hembacher, Lyons, & Ghetti, 2014; Lyons & Ghetti, 2013). Coughlin et al, for example, asked three- to five-year-olds to make metacognitive judgements and control decisions in relation to a perceptual discrimination task. On each trial the children were required to identify a target image from two degraded images. For example, in one trial the children were shown an image of a lemon and an onion with parts of each image removed. The children were then asked to identify which of the two images was the lemon. Immediately after making their decision the participants rated their confidence in their answer. Participants also completed another condition using the same task whereby they could request help before selecting their answer and rating their confidence. From this Coughlin et al found evidence of metacognitive monitoring across all the age groups, with children rating higher confidence for correct answers and lower confidence for incorrect answers. There was also evidence of metacognitive control across all age groups, with participants asking for help on trials that they had previously reported low confidence for.

Interestingly, the age at which children are able to monitor and control their cognitive processes may vary based on the object-level task. Hembacher and Ghetti (2014), for example, found no evidence for metacognitive monitoring or control in relation to a memory recognition task among three-year-olds. In Hembacher and Ghetti’s study three- to five-year-olds were shown a series of images, after which they were presented with pairs of images (one previously seen and one new). The children were then required to identify which of the two images they had previously seen. Immediately after making their decision the children rated how confident, they were that their answer was correct. The children then choose whether to share their answer with the experimenter or to withhold it. Participants were informed that the proportion of correct answers shared with the experimenter would determine the size of their prize at the end. From this, Hembacher and Ghetti found that five-year-olds gave higher confidence ratings for correct answers than incorrect answers and chose to share more correct than incorrect answers with the experimenter. In contrast, three-year-olds did not show any difference in their confidence ratings for correct and incorrect answers and shared a mixture of correct and incorrect answers with the experimenter. Thus, whilst there was evidence of metacognitive monitoring and control in five-year-olds, three-year-olds did not appear to show monitoring or control.

Hembacher and Ghetti findings contrast with Coughlin et al results despite both studies having used the same judgement-of-confidence paradigm. Therefore, the age at which children begin to monitor their own cognitive processes may depend on the object-level task. In this case, three-year-olds appear to be able to monitor their perceptual judgements prior to their memory judgements. These findings highlight the importance of examining metacognitive judgements across a variety of object-level task when investigating metacognitive awareness.

There is relatively little research into metacognition in children younger than this, but some argue that there is evidence of metacognitive awareness from as early as 18 months of age, when infants begin to use mental state language (Bartsch & Wellman, 1995; Wellman & Wooley, 1990). There is also some evidence of uncertainty monitoring in children as young as two and half (Call & Carpenter., 2001; Goupil, Romand-Monnier & Kouider, 2016). For example, Goupil et al. conducted a study with 18-month-olds where the experimenter hid a toy under one of two boxes and asked the infant to point to the location of the toy. In some trials the infant saw where the item had been hidden (possible trials) and in other trials a curtain was drawn before the toy was hidden so that the infant couldn't see (impossible trials). The difficulty of the trials were also manipulated by varying the duration between the toy being hidden and the infant being asked to indicate where the toy was. The results showed that children were able to ask for help when they were unsure of the location (i.e., they would ask for help on the impossible trials). They were also more likely to ask for help when there was a longer delay in identifying the location. These findings suggest that infants as young as one and half are able to respond to various levels of uncertainty.

Importance of Metacognition

From the brief examples already presented, it is clear that metacognition can have a significant impact on our behaviour and everyday functioning. Moreover, the importance of metacognition has been highlighted across various areas of applied research including education, employment, and eyewitness testimony.

Education and Metacognition. Since Brown (1977) first highlighted the importance of metacognition in education, it has been extensively researched. Evidence shows that metacognitive knowledge, monitoring and control all play a central role in academic performance (Tobias & Everson, 2009). Knowing which strategies are most effective for learning, accurately monitoring one's own progress, and adapting one's behaviour accordingly, enables students to perform to the best of their ability. Metacognition has been shown to be central across multiple areas of education, including the sciences, mathematics, and languages (Zohar & Barzilai, 2013; Şahİn & Kendir, 2013; Anderson 2002, 2012). Moreover, metacognition contributes to academic success independent of intelligence (Pishghadam & Khajavy, 2013; Veenman, Kok & Blote, 2005; Veenman & Spaans, 2005; Veenman, Wilhelm, & Beishuizen, 2004). Veenman and Beishuizen (2004), for example, found that in a text comprehension exercise, the amount of variance accounted for by metacognition equalled, and in one task exceeded, the amount of variance accounted for by intelligence.

Teaching metacognitive skills to pupils has also been shown to have a significant impact on pupils' academic outcomes (Baas et al. 2015; Dignath, Buettner, & Langfeldt 2008; Donker et al. 2014; Moely et al., 1992 Perry, Albeg, & Tung 2012; Perry, Lundie & Golder, 2019). A study of 350 students, for example, found that participants who had received metacognitive training were significantly better at solving maths problems in comparison to participants who had not received specific metacognitive training (Fuchs et al., 2003). Overall, research has highlighted the importance of metacognition in education, particularly when it comes to academic success. Therefore, developing a comprehensive understanding of metacognition will have an impact on how metacognition is to be understood within the field of education.

Employment and Metacognition. Given that metacognition is fundamental to one's ability to learn and subsequently achieve academic success, it is not surprising that metacognition has also been noted as a key skill in the world of work. So much so that metacognition, along with skills such as ICT, is now one aspect of what some business sectors refer to as "twenty-first century skills" (Van Laar et al., 2017). It has been incorporated into one of the most influential models of employability, Knight and Yorke's (2002) USEM model, with the 'M' standing for metacognition. This model highlights the importance of metacognition in preparing students for employment. Research has also found that metacognition is important when it comes to co-operation and working successfully within a team (Nonose et al., 2014), and is linked to job performance, career achievement, and effective management and leadership (Clark, 1988; Markessini, 1991; Marshall-Mies et al., 2000; Mikulecky, & Ehlinger, 1986; Mumford et al., 1993; Zaccaro, et al. 1997). Overall, metacognition is becoming seen as an increasingly important aspect of employability (Van Laar et al., 2017) and therefore understanding metacognition may have an impact on obtaining and sustaining employment as well as career development.

Eyewitness Testimony and Metacognition. Another key area of research into metacognition is eyewitness testimony. When eyewitnesses are called upon they are expected to ‘tell the truth, the whole truth and nothing but the truth’. This requires the witness to identify which memories are accurate, and to withhold false ones. Evidence shows that it is easy to plant false memories in one’s mind. For example, Loftus, Miller and Burns (1978) found that by just asking a question that suggests some false information about an event could lead to false memories. Research by Zaragoza and Mitchell (1996) showed that by repeatedly suggesting the same piece of false information participants would report higher confidence in their memory despite it being incorrect. This clearly demonstrates how fallible one’s metacognitive accuracy can be. This is a particular issue when it comes to eyewitness reports considering that eyewitness’ who show they are confident in their reports are more likely to be believed, (Cutler, Penrod, & Dexter, 1990; Fox & Walters, 1986; Loftus & Ketcham, 1991). Therefore, gaining a full understanding of metacognition can shed light on the benefits and potential pitfalls of relying on eyewitness testimony in court.

Overall, it is clear that metacognition is important for everyday functioning as well as academic success, employment and criminal justice. Thus, it is important to study metacognition from an applied perspective, however, it is also important to study metacognition from a theoretical perspective. Metacognition is thought by many to be closely related to the ability to represent mental states in others, or to “mindread” (also known as metalizing or theory of mind). Indeed, as originally defined, theory of mind is the ability to represent mental states, such as beliefs, desires, and intentions, to oneself *and* others in order to explain and predict behaviour (Premack & Woodruff, 1978). Inherent in this original definition was the view that representation of one's *own* mental states and cognitive activity relied on the same set of mechanisms/processes as the awareness of *others*' mental states and cognitive activity. Although the study of metacognition has diverged from the study of mindreading, theories of the relation between the two have been built and pitted against each other over the past thirty years.

Theories of Mindreading and Metacognition

Given that in Premack and Woodruff's (1978) original definition, theory of mind encompasses the ability to attribute mental states to oneself and others, it is not surprising that a substantial amount of attention has been given to the relationship between these two abilities. In recent decades several opposing accounts have emerged to explain the relationship between these two abilities, most notable of these are the theory-theory (one-system/mindreading-is-prior; Gopnik, 1990; Perner, 1991; Carruthers, 2009), simulation theory (metacognition-is-prior; Goldman, 2006), and the two-system theory (monitoring mechanism theory; Nichols & Stich, 2003).

Theory-Theory

Over the past 30 years the theory-theory account of meta-representation has received a significant amount of attention. This account argues that people have a 'theory of mind' which enables them to attribute mental states, such as beliefs, judgements, intentions and desires, to themselves and others. The 'theory of mind' provides a set of guiding principles regarding how these mental states interact. For example, people act in accordance with their beliefs, thus if Sophie tells you she wants her toy and then goes to the toy box, you may conclude that Sophie must *believe* that her toy is in the toy box. How ‘theory of mind’ develops is still up for debate with some arguing that children develop a ‘theory of mind’ through experience, and thus the child acts like a scientist, learning how the mind works (Gopnik, 1992). Others claim that humans possess an innate ‘theory of mind’ mechanism that matures with age (Carruthers, 2009; Leslie, 1987).

Irrespective of its development, and in line with the heuristic-based theory of metacognition, proponents of theory-theory argue that people do not have direct-access to their own metal states. It is argued that people attribute mental states to themselves using the same mechanism, or theory of mind, as that used for attributing mental states to others. Essentially the theory-theory account claims that metacognition relies on the same inferential, metarepresentational processes as mindreading. Hence, proponents of this theory state that metacognition is the result of turning our mindreading abilities in on ourselves and is therefore the results of unconscious self-interpretation.

As with how ‘theory of mind’ develops, there is also disagreement among proponents of the theory-theory account in terms of the specific input available to self-interpretation. Gopnik (1993), for example, argued that self-interpretation is based on observation of overt behaviour and physical circumstances. This theory, however, came under heavy criticism by proponents of simulation theory (Goldman, 2006) and the two-system theory (Nichols & Stich, 2003). It was highlighted that one can become aware of some internal state, such as having decided to put an offer in on a house they viewed earlier in the week, despite sitting absolutely still. Carruthers (2009), a strong advocate for theory-theory, concedes that self-interpretation cannot purely rely on physical circumstances and overt behaviour but that additional information such as inner speech, imagined images, attentional focus and emotional input must also inform self-interpretation.

Regardless of the variation in how the theory of mind mechanism develops and what inputs are available for self-interpretation, the various theory-theory accounts converge on the fundamental principle that humans must infer their own mental states using the same metarepresentational processes as those used to infer others mental states. Thus, the theory-theory account predicts that there should be no dissociation between the two capacities. Therefore, if a person has difficulties with mindreading, they will also have difficulties with metacognition, and likewise, if metacognition is diminished mindreading will also be diminished. This prediction is fundamental to the studies presented in this thesis and contrasts with the predictions made by both the simulation theory and the two-system theory.

Simulation Theory

In contrast to theory-theory but in line with the direct access account of metacognition, simulation theory argues that people have direct access to their own mental states via introspection. That is to say that there is no need to infer our own mental states based on perceptual input, rather we can just know our own mental states. Key to the mindreading and metacognition debate is that proponents of simulation theory argue that we infer others' mental states via mental simulation of our own directly accessible metal states (Goldman, 2006). In other words, people adopt the perspective of the other person and reflect on what their own mental states would be in that situation (introspect), and then attribute those mental states to the target person. This contrasts with the theory-theory account in the sense that mindreading relies on metacognition. Thus, if metacognition is diminished then so too is mindreading. Equally, it is possible that, in cases where the simulation component is impaired, one may be able to introspect accurately but have difficulty with mindreading. Therefore, simulation theory predicts that there should be cases where mindreading and metacognition are dissociated. Again, this prediction is key to the studies presented within this thesis and will be discussed in more detail, below.

Two-System Theory

Finally, the two-system theory, in line with simulation theory and direct-access theory, claims that people have privileged first-person, non-inferential access to their own mental states. In contrast to simulation theory, however, the two-system account asserts that humans possess an innate, specialised mechanism for 'detecting' their *own* metal states, which is distinct from the processes used to detect *others* mental states (Nichols & Stich, 2003). Nichols and Stich define 'detecting' as the capacity to attribute mental states to oneself or others (e.g. I/he believes that x). They argue that the innate monitoring mechanism comes 'online' early in development, and before the ability to 'reason' about mental states. 'Reasoning', being the ability to predict behaviour in accordance with mental state attribution. Nichols and Stich argue that reasoning about one's own mental state depends on a 'theory of mind', similar to that described by theory-theorists. Moreover, they argue that the same theory of mind process is used to reason about both one's own mental states and others. Thus, the ability to attribute mental states to self and others relies on distinct mechanisms but the ability to reason about one's own and others mental states relies on the same processes.

At the heart of the two-system theory is that *others'* mental states must always be inferred on the basis of behaviour, circumstance and knowledge of the target person, whereas one's *own* mental states are detected by the 'special monitoring mechanism' and therefore no inference is required. Thus, this account predicts that it should be possible to find examples where mindreading is diminished but metacognition is intact and vice versa. Again, this will be explored in more detail throughout this thesis.

Mindreading

Development of Mindreading

Mindreading is a highly researched area in psychology and the majority of this research has focused on the *development* of mindreading abilities in children. In particular it has focused on false beliefs, the awareness that another person can hold a belief that does not match reality. The classic tests of the awareness of false beliefs include the unexpected transfer task by Wimmer and Perner (1983) and the Sally-Anne task developed by Baron-Cohen, Leslie, and Frith (1985). Both tasks involve the participant listening to a story about a protagonist placing an item in a specific location (location A) and then leaving the room. Whilst the protagonist is away another character comes in and moves the item to a different location (location B). The participant is then asked where they expect the protagonist to look. The correct answer is location A. By declaring that the protagonist will look in 'location A' the participant is showing that they are able to distinguish between reality (where the item actually is) and how others may (incorrectly) represent the world (where the protagonist believes the item is) and hence are said to possess mindreading abilities.

Many studies have shown that typically developing children are able to pass false belief tasks at around the age of four (Wellman, Cross, & Watson, 2001). This has led to the conclusion that children do not develop mindreading abilities until the age of four

(Baron-Cohen, et al., 1985). Since the development of these ground-breaking tasks research has challenged the idea that children younger than four lack a theory of mind. It has been argued that younger children's difficulties on these traditional false belief tasks may be to do with the cognitive demands of the task itself, rather than lacking mindreading abilities (Baillargeon et al., 2010; Leslie & Polizzi, 1998). Research using eye gaze and looking time has suggest that even infants show some level of mindreading. For example, Clements and Perner, (1994) found that whilst 35-month-old infants failed the verbal aspect of a false belief task they did look towards the 'correct location', that is where the protagonist believed the item was and not where the item actually was. This is further supported by tasks that involve the violation of expectations. Onishi and Baillargeon (2005) for example found that 15-month-olds looked longer at a scene if the protagonist reached for where the item actually was rather than where they had left it. Research has also found similar results in 7-month-olds (Kovács et al., 2010). There is, however, some debate over what this data implies. Perner and Ruffman (2005), argue that this type of data merely reflects that the infant has learnt 'behavior rules' (i.e people look for things where they last saw them). Others have suggested that it may imply implicit mindreading but that it does not provide evidence for 'fully fledged' mindreading (Perner & Roessler, 2012). More, recent research has attempted to address some of the debates. For example, Buttleman et al. (2014) found that 18-month-olds were able to act on their knowledge in line with what would be expected if they thought the protagonist had a false belief. To act on one's awareness implies that their knowledge must be more than just implicit (Perner, 2014).

This debate, like the debate concerning the development of metacognition is likely to continue for some time yet (See Philips et al 2015 and Kaddouri et al, 2020). Crucially though, for the debate concerning the relationship between mindreading and metacognition there does appear to be some parallel in the development of mindreading and metacognition. Take the unexpected content task for example. In this task participants are shown a smarties tube and asked what they think it contains. The correct answer is of course ‘Smarties’. The researcher then opens the tube and shows the participant that the tube is in fact filled with pencils. The participant is then asked a) what they thought was in the tube before they were shown it (self-test) and b) what someone else, who has not seen the actual contents, may think is in the tube (other-person test). To pass the ‘self-test’. participants must indicate that they initially thought that the tube contained smarties. Likewise, to pass the ‘other-test’, participants must say that someone else, who has not seen the true content, will think that it contains smarties. Research shows that typically developing children pass the self and other test questions at the age of four, indicating that mindreading and metacognition develop in parallel (Wellman et al., 2001).

Advanced Mindreading

Mindreading can be measured in numerous ways and a plethora of tools have been developed to test more advanced mindreading abilities beyond the tasks previously described. For example, more complex second order false beliefs are often measured using the ‘strange stories’ task developed by Happé (1994). In this task participants read 24 short vignettes and are asked to explain why one of the characters says something that is not literally true. To show mindreading abilities, the participant is required to attribute mental states such as beliefs, intentions and desires to the character and in some cases identify what one character may believe about another character’s metal state (e.g. Sophie thinks Jim thinks…).

Other measures of more advanced mindreading abilities include Reading the Mind in the Eyes (RMIE; Baron-Cohen et al. 2001), the Animation task (Abell et al. 2000), and the Movie for the Assessment of Social Cognition (MASC; Dziobek et al. 2006). In the RMIE task participants are shown 36 photographs of eyes and are required to select what the person in the picture is feeling, out of four possible emotions. This task differs from other measures in the sense that it only requires the attribution of one type of mental state, namely emotions. It also places less demand on the participant to integrate situational details, allowing the participant to focus purely on metal state attribution (Chung, Barch & Strube, 2011). Considering the narrow nature of this task it has been argued that it is a measure of emotion recognition rather than mindreading per se (Oakley et al. 2016; Olderbak et al. 2015), although see Nicholson et al, (2019) who highlight evidence which shows that RMIE correlates with other measures of mindreading (Jones et al., 2018) as well as activating the same brain regions that are activated when completing other mindreading tasks (Schurz et al., 2014). To overcome this potential criticism other measures can be used to complement the use of the RMIE. The Animation task and the MASC, for example, provide a measure of spontaneous mental state attribution. In these tasks participants watch videos and attribute mental states, such as beliefs, desires, and intentions. In the Animation task participants watch four short video clips of triangles moving around a screen, and then describe what they believe to be happening in the clip. The MASC on the other hand requires participants to watch a film depicting a group of people interacting. The film is stopped at regular intervals and the participant is given a multiple-choice question regarding the mental states of a particular character at the moment the film was stopped. In both cases mindreading ability is assessed based on the accuracy of mental state attributions. Both these are more complex than the RMIE, in the sense that they require the attribution of more than one type of mental states and also place a higher demand on other cognitive capacities.

Linking Mindreading and Metacognition: The Case of Autism Spectrum Disorder

Evidence for the relationship between mindreading and metacognition has been drawn from a wide variety of sources and domains of psychology (Carruthers, 2009; Goldman, 2006; Nichols & Stich, 2003). Perhaps the most significant evidence has come from studies of one particular developmental disorder, namely autism spectrum disorder (ASD). Autism is a neurodevelopmental condition characterised by atypical social-communication and restricted/repetitive behaviours and interests (American Psychiatric Association, 2013). According to the World Health Organisation (2021) it is estimated that 1 in 160 children across the world are autistic[1], although some research produces figures much higher than this (Baird et al., 2006; Elsabbagh et al., 2012; Kawamura, Takahashi, & Ishii, 2008; Kim et al., 2011). There is also evidence of milder or subclinical characteristics of autism among the general population (Baron-Cohen et al., 2001; Constantino & Todd, 2003). These milder or subclinical characteristics found within the general population are often referred to as the broad autism phenotype (BAP).

[1] The author acknowledges that there is an ongoing debate among individuals who have been diagnosed with ASD concerning the terminology used when referring to individuals with such a diagnosis. This thesis uses ‘autistic individuals’, however the author acknowledges that some individuals diagnosed with autism prefer ‘individuals with autism’. Due to the ongoing debate, the author has chosen to use ‘autistic individuals’’ in line with recent research highlighting a preference for identify first language, over person-first language, among those diagnosed with autism and family members (Fletcher-Watson et al., 2019; Kenny et al., 2016). However, it is acknowledged that ‘individual with autism’ may be preferable to some people and therefore the author does not mean to cause any offence with the terminology used within this article.

In addition to the complexities of autism, autistic individuals often have high rates of comorbid disorders including, but not exclusively, attention-deficit-hyperactivity-disorder, alexithymia, anxiety, specific language impairment, and intellectual disability (Simonoff, Pickles, Charman, Chandler, Loucas & Baird, 2008; Nebel-Schwalm, & Worley, 2014). The high level of comorbidity has made it particularly difficult to gain a full understanding of what causes autism. Thus, no theory, genetic, cognitive, or otherwise appears to be able to explain all aspects of autism. Twin and family studies, for example, have provided strong evidence that autism has a genetic basis with high heritability (Bailey et al. 1995; Bolton et al 1994; Folstein and Rutter 1977; Hallmayer et al., 2011; Le Couteur et al., 1996). However, a vast array of genes have been implicated in the various aspects of autism, some of which overlap with co-occurring conditions (Happé & Ronald, 2008). Likewise, at a cognitive-level many theories have been proposed but not one theory is able to account for all aspects of autism (Happé & Ronald, 2008). Despite this, the 'weak central coherence' theory (Frith & Happé, 1994; Happé & Frith, 2006), the 'executive dysfunction' theory (Steel, Gorman & Flexman, 1984; Rumsey, 1985) and the 'theory of mind' theory (Baron-Cohen, Leslie & Frith, 1985) have gained the most attention over the years.

Weak central coherence

The weak central coherence account explains autism in terms of a tendency to process information at a local level rather than a global level (e.g., focusing on detail/parts rather than incorporating perceived information to obtain the global gist of it). Evidence supporting this theory comes from research showing that autistic participants are better than typically developing participants at tasks that require local level rather than a global level processing such as the embedded figures tasks and block design tasks (Shah & Frith, 1983). The embedded figure task, for example, requires participants to identify a simple shape within a more complex image. Research shows that autistic participants are more accurate and quicker when it comes to this task (White & Saldana, 2011). Further evidence comes from face processing studies that find that autistic people do not appear to show the inverted face effect, that is, their facial recognition is not affected when faces are presented upside down (Langdell, 1978). From this, and other studies manipulating the attentional focus during face processing it was concluded that autistic people process the individual features of faces rather than the face as a whole (Lopez, Donnelly, Hadwin, & Leekam, 2004). Despite this atypical processing, and its link to the non-social aspects of autism, such as restricted and repetitive behaviours, it fails to account for the full complexity of the social-communication aspect of autism, and therefore an alternative theory in required (Happé & Frith, 2006).

Theory of mind

Arguably the most influential theory of autism to date, is the theory of mind/mentalising theory (Baron-Cohen et al., 1985). According to this theory, the main behavioural features of autism, or the social-communication difficulties at least, are caused by a cognitive-level difficulty with inferring the mental states of others (mindreading). In keeping with this theory, autistic children tend to show diminished performance on traditional mindreading tasks, compared to age- and ability-matched non-autistic individuals (Happé, 1995). This difficulty with mindreading has also been shown to continue into adulthood, across a number of age-appropriate mindreading tasks (Baron-Cohen et al. 2001; Brewer, Young & Barnett, 2017; Grainger, Williams & Lind, 2014; Happé 1994;

Lind, Williams, Bowler & Peel, 2014). This between-group difference in mindreading ability is usually large in magnitude (e.g., Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998), though there is some evidence to indicate that the difference is shrinking over time (e.g., Rødgaard, Jensen, Vergnes, Soulières, Mottron, 2019). Importantly, several (though not all) studies show a negative association between mindreading ability and autism feature severity (as mindreading ability decreases, so severity of autism features increases (see Brunsdon & Happé, 2014). The theory of mind account has been invaluable at explaining the social features of autism, however, it struggles to account for the non-social aspects.

Executive Dysfunction.

An alternative theory that has attempted to explain the social and non-social aspects of autism is the executive dysfunction account. Executive function refers to a range of cognitive factors including inhibition, cognitive flexibility, and working memory (Miyake et al., 2000), all of which are implicated in a variety of cognitive processes such as planning and decision making. In line with this theory, research indicates that planning is impaired among autistic people as demonstrated by difficulties in the classic Tower of Hanoi task (Bennetto, Pennington, & Rogers, 1996; Ozonoff & Jensen, 1999; Ozonoff & McEvoy, 1994; Ozonoff, Pennington, & Rogers, 1991). The tower of Hanoi task requires participants to move disks arranged on three different pegs to an arrangement designated by the experimenter, following a set of rules and in as few moves as possible. Cognitive flexibility has also been shown to be impaired among autistic people, with difficulties on tasks such as the Wisconsin Card Sorting task (Ozonoff & Jensen, 1999; Prior & Hoffmann, 1990; Ozonoff & McEvoy, 1994). The Wisconsin Card Sorting task requires participants to sort cards in line with an unspoken rule which the experimenter can change at any point. The key to success on this task is to adapt ones sorting in line with the feedback of correct/incorrect sorting provided by the experimenter.

In contrast to the evidence on planning and cognitive flexibility, autistic people do not seem to be impaired on tasks that employ inhibition such as the Stroop test and the Go/no-go task (Ozonoff & Jensen, 1999; Ozonoff & Strayer, 1997; Ozonoff, Strayer, McMahon, & Filloux, 1994). Each of these tasks require participants to inhibit a response, for example, in the Stroop test participants are presented with the name of a colour (RED) with the font in the corresponding colour (RED) or a different colour (PURPLE). As each word is presented, participants are required to read what the word says rather than saying the colour of the font. This takes advantage of the fact that people have a bias towards stating the colour they see rather than what the actual word says and therefor they have to inhibit their initial response.

The executive dysfunction theory can account for restricted and repetitive behaviours and to some extent the impairments found in mindreading tasks (Frye et al., 1995; Ozonoff et al., 1991). Some, for example, argue that mindreading tasks require participants to inhibit their own knowledge/beliefs about their reality in order to infer others mental states. Thus, people may fail false belief tasks due to executive dysfunction rather than a lack of theory of mind. However, one significant challenge to this account is that deficits in executive functions are not universal among autistic people (Hill, 2004). Furthermore, there appears to be little evidence for executive dysfunction in pre-school children, suggesting that it may be a secondary deficit in autism that only develops later on (Hill, 2004).

Fractionable Account.

More recently it has been suggested that autism is ‘fractionable’ at the cognitive level, and therefore no single factor can explain all aspects of autism (Brunsdon & Happé, 2014; Happé & Ronald, 2008; but see Hobson, 2014, and Mottron 2021). The theory of mind hypothesis provides a good explanation for atypical social-communication but struggles to account for the non-social aspects of autism. The weak central coherence theory can account for the non-social aspects but fails to account for the social aspects. Finally, the executive dysfunction account can potentially explain both the social and non-social aspects but difficulties in executive function are by no means universal among autistic people (Hill, 2004). Thus, the factional account of autism posits that distinct behavioural aspects of autism are influenced by difference cognitive components.

Metacognition in Autism

Given the abundance of evidence indicating that mindreading is diminished in autism it is not surprising that psychologists and philosophers have been drawn towards autistic people to provide evidence clarifying the relationship between mindreading and metacognition. After all, each theory makes its own predictions regarding the state of metacognition if mindreading is diminished (see Table 1). Theory-theory predicts that if mindreading is diminished then metacognition will also be diminished. In contrast the two-system approach and simulation theory both predict that it is possible for metacognition to be intact despite mindreading difficulties. Proponents of all three of these theories have drawn on autism research to support their hypothesis. The two system-approach and simulation theory, for example, claim that metacognition is intact among autistic people and therefore provides evidence in support of their theories. The one-system approach, on the other hand claims, that metacognition is diminished among autistic people.

Table 1

Predictions if Mindreading is Diminished

Theory	Mindreading	Metacognition
Theory-Theory	X	X
Simulation Theory	X	✓
Two-System	X	✓

Note. X = Diminished; ✓ = Intact

Initial evidence regarding the metacognitive competency of autistic people drew upon autobiographical and introspection data. Nichols and Stich, who advocate for the two-system account, claim that a study conducted by Hurlburt, Happé and Frith (1994) provides clear evidence for intact metacognition. In the study three adult males with a diagnosis of Asperger Syndrome were given a device and asked to write down their inner experiences every time the device beeped. Each participant was also interviewed about each recorded experience. Nichols and Stich take extracts from Hurlburt et al's (1994) study highlighting that the participants were reporting their current thoughts and feelings. They argue that this provides evidence for the notion the autistic individuals have intact metacognition despite having diminished mindreading abilities.

One significant critique of Nichols and Stich's interpretation of Hurlburt et al's study is that all three men were able to pass basic first level-false belief tasks (Carruthers, 2011), such as the Sally-Anne task described previously. Two of the men were also able to pass more complex second order false belief tasks, where participants are required to ascribe beliefs to one character about another character's beliefs (Perner & Wimmer, 1985). The fact that each of the participants were able to successfully complete these measures of mindreading indicate that they all had some mindreading capabilities. Therefore, this does necessarily not show a dissociation between mindreading and metacognition as predicted by Nichols and Stich. After all, you would expect some level of metacognition from individuals who pass false belief tasks. Interestingly and equally as challenging to Nichols and Stich's interpretation of Hurlburt et al's study, is that there was a strong correlation between the individual's performance on the false belief tasks and the sophistication and ease of their introspective reports (Carruthers, 2009). This is exactly what would be expected if mindreading and metacognition relied on the same processes, and therefor challenges the notion that there is a specialised mechanism for 'detecting' one's own mental state that is distinct from the mechanism(s) for detecting others mental states.

Further evidence presented by Nichols and Stich (2003) in support of the two-system account comes from autobiographical accounts of autistic adults regarding childhood memories (Frith & Happé, 1999). They claim that the fact that the autistic adults were able to recall events from their childhood clearly demonstrated an awareness of their own mental states. Nevertheless, as Carruthers (2011) points out, there is no reason to suspect that memory formation is a metarepresentational process or that metarepresentation occurred at the time of memory formation, rather it just shows that the individuals reporting those autobiographical accounts are capable of forming memories.

Following the debate concerning the interpretation of the introspection and autobiographical data, Williams (2010) presents evidence from studies examining own false beliefs and intentions (Baron-Cohen, 1991, 1992; Fisher, Happé & Dunn, 2005; Russell & Hill, 2001; Phillips et al., 1998; Williams and Happé, 2009a; Williams & Happé, 2010), concluding that metacognition is impaired among autistic people. For example, Williams draws on evidence from the unexpected content (smarties) task, previously described. Research using this type of unexpected content task shows that autistic participants find both the ‘self ‘and ‘other’ question significantly *more* difficult than typically developing children.

A major criticism of the evidence presented by Williams is that it relies on participants reporting past mental states rather than current metal states (Carruthers, 2009; Grainger, Williams & Lind, 2014). This is a particular issue for the theoretical debate on the relationship between mindreading and metacognition because simulation theory and the two-system account argue that it is *current* mental states that are directly accessible without the need for mindreading. Therefore, if the theoretical debate is to be addressed more precisely research must use methods that examine *current* mental states such as judgements-of-confidence, feeling-of-knowing, judgement-of-learning and post-decision wagering.

Current Thesis

Considering the impact that examining metacognition among autistic people can have from both a theoretical and clinical perspective, it is crucial that we first get a clear picture of the existing research in this field. Therefore, the next chapter (chapter two) in this thesis will provide a meta-analysis and critical review of the existing research that has employed online methods to assess metacognition in autism. As noted previously, examining current or online mental states is central to the theoretical debate, hence focusing research that employs online metacognitive tasks. This is the first meta-analysis into online metacognition in autism, and so provides a novel overview of the research in this field, highlighting gaps and areas for future research. The remaining chapters will heed the recommendations from the meta-analysis and critical review and build upon the existing research using original studies to understand metacognitive processing in autism. Each study examines mindreading, metacognition and autism using a combination of case-control and individual differences experiments.

The first study, chapter three, examines metacognitive awareness using a post-decision wagering paradigm. It examined both accuracy and ease with which meta-level judgements are made. Experiment 1 examines these variables in relation to each other in a student population. Experiment 2 compares metacognitive accuracy and reaction time in an autistic sample and a typically developing sample. This is the first study to examine metacognitive accuracy in autistic people using post-decision wagering, and the first to employ reaction time data to examine the ease at which meta-level judgements are made. Thus, this study provides a novel look at metacognitive accuracy in autism on several counts. Chapter four extends these findings by using the same object-level task as chapter three but uses judgements-of-confidence as a measure of metacognitive awareness. This is the first time that the ease at which judgements-of-confidence are made have been examined in autistic people. As with chapter three, this study presented 2 experiments, one using an individual differences approach and one using a case-control approach.

Following on from this, chapters five and six investigates the landscape of metacognition in autism by examining the extent to which metacognition is globally impaired. Chapter five presents a study investigating prospective as well as retrospective judgments-of-confidence within a common sample. No study has yet investigated prospective judgments-of-confidence in autism, and therefore, this provides a new opportunity to gain an understanding of past and future meta-cognitive judgements in autistic people. Chapter six employs the same post-decision wagering paradigm that was used in chapter three but manipulates the object-level task to compare meta-level awareness of perceptual and semantic decisions among a common sample of autistic participants. Again, this is the first time that any study has examined perceptual and semantic metacognition using post-decision wagering within a common sample.

Chapter seven moves from looking at judgements-of-confidence and post-decision wagering, to investigate metacognition using a feeling-of-knowing and remember-know-guess paradigm, both of which have received very little attention in autism research. Finally, chapter eight will amalgamate the findings from each of these studies along with the existing research to produce an overview of metacognition in relation to autism and mindreading, with some recommendations for future research, as well highlighting what these findings mean from a theoretical and clinical perspective.

It should be noted that the autistic and typically developing participants who participated in the studies presented in this thesis were all recruited from the Autism Research at Kent database and therefore several participants took part in multiple studies. In total 103 participants took part in the studies presented in this thesis, 50 autistic participants and 53 typically developing. Nine (4 autistic) of the participants took part in all five studies, five (3 autistic) took part in four out of the five studies, 24 (12 autistic) took part in three out of the five studies, 40 (20 autistic) number took part in two out of the five studies, leaving 25 (11 autistic) that only participated in one study.

Chapter Two: A Meta-Analysis and Critical Review of Metacognitive Accuracy in Autism

Abstract

This chapter presents a meta-analysis and critical review of the existing literature examining metacognition in autism. The aim of this meta-analysis was to establish if metacognition is impaired among autistic people and, if so, to what extent. A total of 17 studies comparing 412 individuals diagnosed with autism and 453 typically developing individuals were included in the meta-analysis. The data revealed a moderate, albeit heterogenous, impairment in metacognitive accuracy among autistic individuals in comparison to typically developing individuals. A critical review of the results showed that whilst there may be an overall deficit, it is important to take into account the age group being examined (children or adults), as well as the meta-level task being employed when drawing conclusions about autistic individuals' metacognitive abilities.

Metacognition refers to cognitions about our own mental states (Flavell, 1979). It is crucial for how we live our lives, and underpins how we make sense of, predict, and control our actions. It plays a key role in learning and decision making and predicts academic performance independently of general intelligence (Dunlosky & Metcalfe, 2009). Therefore, difficulties with metacognition are likely to have significant implications for everyday functioning. As such, understanding metacognition and its processes is key to supporting individuals with diminished metacognitive abilities.

Metacognition can be divided into three key components: metacognitive knowledge, metacognitive monitoring, and metacognitive skills (Flavell, 1979). All these aspects are important for everyday life, but the majority of research into metacognition has focussed on meta-monitoring, in part because metacognitive skills depend on monitoring. Metacognitive monitoring is how we represent occurrent, ongoing cognitive activity, such as judging how confident we are that we have learned all we need to pass an exam. There is also a rich history of theorising about a potential link between metacognitive monitoring and mindreading (also known as mentalizing), which is the ability to represent the mental states of others. Some theories predicted that if mindreading is impaired then metacognition will also be impaired whereas others predict that it is possible to have intact metacognition despite having impaired mindreading (Carruthers, 2009; Goldman, 2006; Nichols & Stich, 2003).

These clinical and theoretical issues make it important to study metacognition in neurodevelopmental conditions, especially in conditions that involve difficulties with mindreading, and arguably, the condition most clearly associated with mindreading difficulties is autism spectrum disorder (ASD). Autism is a developmental condition diagnosed on the basis of restricted and repetitive behaviours, and social-communication difficulties (American Psychiatric Association 2013; World Health Organisation 2018).

Individuals diagnosed with autism have been shown to have difficulty representing mental states in others (Happé & Frith, 1995). There is now emerging evidence suggesting that individuals diagnosed with autism also struggle to represent their *own* mental states. If this is the case, then poor metacognitive ability may be having a significant impact on numerous aspects of those individuals' everyday lives.

Multiple methods have been used to examine online metacognitive accuracy, including judgements-of-confidence, judgements-of-learning, and feelings-of-knowing. Each of these tasks require the participant to make a judgement about an ongoing mental state. In judgement-of-confidence tasks, for example, participants first make a cognitive/object-level judgement, such as choosing which of two images has the most dots, or answering a general knowledge question. Participants then make a meta-level judgment by rating how confident they are that their object-level answer was correct. The stronger the relationship between the object-level performance and the meta-level judgment, the better the metacognitive accuracy (i.e., high confidence for correct answers, low confidence of incorrect answers). This is often calculated using a Gamma correlation (Kruskal & Goodman, 1954), which is a non-parametric measure of association that indexes the extent to which cognitive-level performance is associated with meta-level judgements. The higher the Gamma score, the better the metacognitive accuracy (ranging from -1 to +1).

Similarly, in judgements-of-learning tasks participants are asked to learn something (e.g., a list of words) and then rate how likely it is that they will remember what they have learnt when subsequently tested (the meta-level judgement). Participants then complete the object-level task of recalling the previously learnt information. Again, the closer the correspondence between meta-level judgement-of-learning and actual learning, the better metacognition is. In feeling-of-knowing tasks, participants are required to indicate how likely it is that they will know the answer to a specific question or recall/recognise a specific piece of information. For example, participants may be asked to memorise a list of word pairs (a cue and a target word) and then, after a distractor task, they will be asked to recall the original target word when presented with the cue word. Following this recall phase, participants are required to make a feeling-of-knowing judgement by indicating if they would recognise the target word when it is presented among four other words. Once the participants have made their metacognitive judgement, they complete the recognition phase/object-level task, whereby they are asked to identify the target word amongst some lure words. As with the judgements-of-learning and judgements-of-confidence tasks, the closer the correspondence between meta-level judgements (feeling-of-knowing judgement) and object-level performance (actual recall) the more accurate metacognition is. There are many variations of each task but key to each of them is that they all assess the participant's current mental state and rely on explicit verbalisation of such judgements.

Given that the debate concerning the relationship between mindreading and metacognition relies heavily on the understanding of current mental states, this meta-analysis will focus on research that examines explicit verbal online metacognitive accuracy among individuals diagnosed with autism in comparison to typically developing (TD) individuals. Due to the variation in such research, this paper will present the initial meta-analysis followed by a critical review of the research included in the meta-analysis. The critical review will take account of key issues when examining such research, reflecting on how such issues relate to the findings of the meta-analysis.

Method

Eligibility criteria

The following eligibility criteria was set out prior to conducting the literature search. To be eligible the studies must have examined individuals of any age (children and/or adults) diagnosed with autism in comparison to a typically developing group. The tasks within the studies had to involve online explicit metacognitive judgements as described above. It was also crucial that the tasks did not involve any aspect that could result in improved metacognitive performance, such as training. Articles were excluded if they did not fit these criteria, were written in a language other than English, did not provide novel data, or did not provide sufficient quantitative data to calculate effect sizes in the form of Hedges' g (e.g., means, standard deviations, p-values, t-tests).

Database Search

A literature search (see Figure 4) was conducted using Web of Science, PubMed and PsychInfo using the search terms "autism" AND "metacognition" for all articles published prior to April 2021 resulting in a total of 675 articles (Web of science = 83; PubMed = 84; PsychInfo = 508). Out of these, 106 were duplicates, 31 were in a language other than English, and 501 were excluded because they either examined something other than metacognition in autism in comparison to a typically developing sample, or they were reviews that did not provide any novel data of their own. Of the remaining 37, 15 used a questionnaire to measure meta-cognition, one examined metacognitive knowledge rather than accuracy (Farrant, Blades, & Boucher, 1999a), one examined metacognitive control (Farrant, Boucher, & Blades, 1999b), and one examined metacognitive accuracy using non-verbal measures (Carpenter, Williams, & Nicholson, 2019). Five articles did not provide the data required to calculate an effect size. The authors of the current meta-analysis attempted to contact the corresponding authors for each of these studies, two of whom provided the data required (Doenyas, Mutluer, Genç, & Balcı, 2019; Maras, Gamble, & Brosnan, 2019). It was not possible to make contact with the corresponding authors of the remaining three studies and therefore they could not be included in the current meta-analysis (Zalla et al., 2015; Brosnan et al., 2016; Wilkinson et al., 2010). This left 16 articles that examined online explicit verbal metacognitive accuracy among individuals diagnosed with autism. There was an additional study that did not come up in the literature search, but the authors of the current meta-analysis were aware of (Wojcik et al., 2011). This resulted in a final sample of 17 independent studies, see Table 2.

Figure 4

Flow-chart depicting literature search process

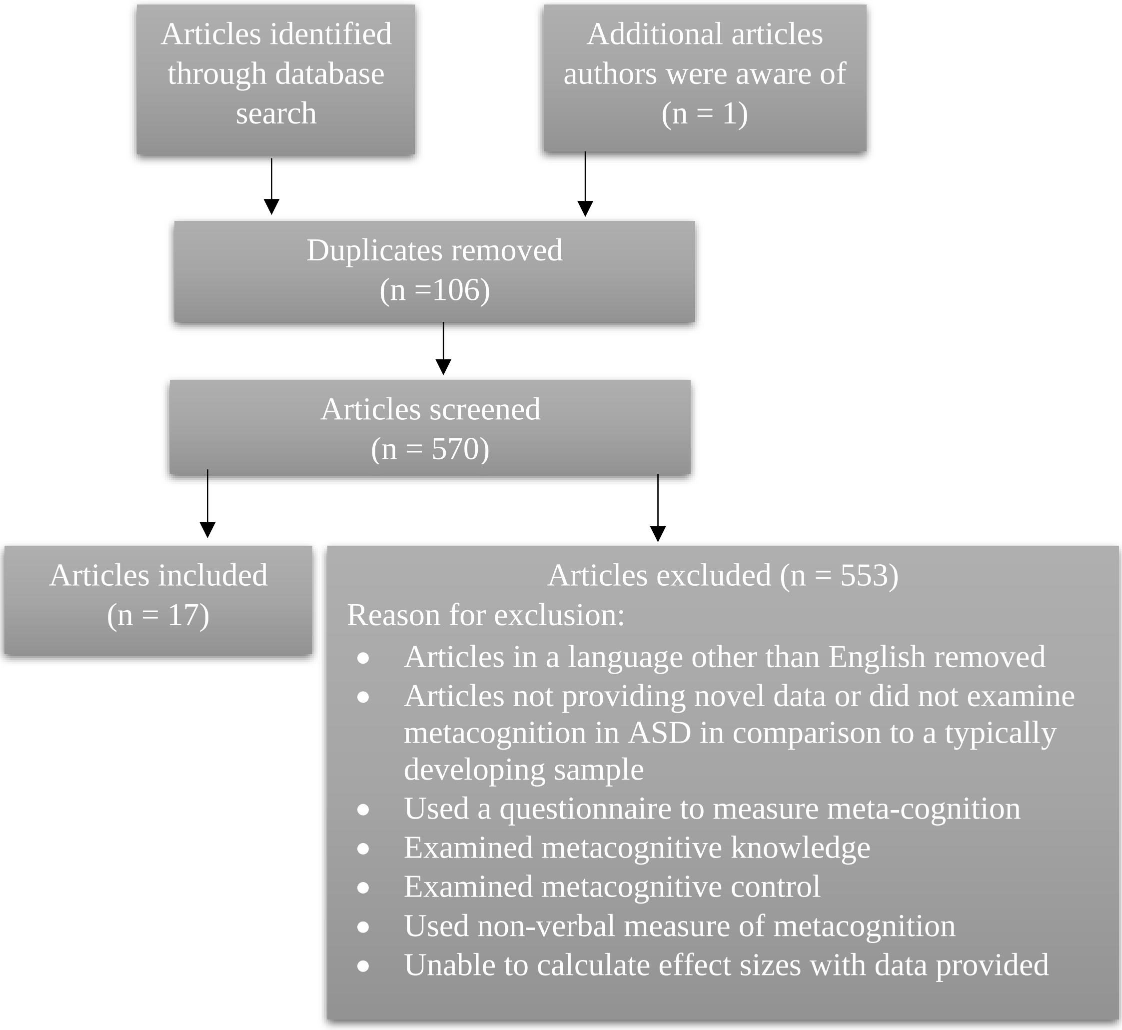


Table 2

Overview of Studies of Metacognitive Accuracy in Autism

Category	Lead Author (year)	Participant Characteristics					Object-level	Hedges' *g*
		Sample size (males)		Mean age (std)		Matched	Matched	
		ASD	TD	ASD	TD			
Judgement-of-confidence								
Adult								
	Cooper et al. (2016)	24 (11)	24 (11)	31.38 (7.28)	30.46 (6.95)	Yes	Yes	-0.25[a]
	Maras, et al. (2020) Social Version	18 (NR)	15 (NR)	34.53 (14.55)	33.67 (10.93)	Partially	Yes	-0.44 [a]
	Maras, et al. (2020) Online Version	18 (NR)	15 (NR)	31.33 (12.14)	35.93 (12.63)	Yes	Yes	0.38 [a]
	Nicholson, et al. (2019)	22 (15)	22 (16)	37.15 (19.94)	37.21 (12.34)	Yes	Yes	-0.66
	Sawyer, et al. (2014)	30 (NR)	52 (NR)	29.9 (11.5)	24.8 (8.4)	Partially	Yes	-0.30 [a]

Category	Lead Author (year)	Participant Characteristics					Object-level	Hedges' *g*
		Sample size (males)		Mean age (std)		Matched	Matched	
		ASD	TD	ASD	TD			
Judgement-of-confidence								
Children/adolescent								
	Doenyas et al., 2019	8 (7)	8 (7)	14.13 (NR)	14.13 (NR)	Partially	Yes	-0.43 [a]
	Grainger et al. (2016a)	32 (NR)	30 (NR)	13.59 (1.36)	13.27 (1.06)	Partially	Yes	-0.41 [a]
	Maras et al., (2019)	16 (12)	46 (28)	13.19 (1.42)	13.43 (1.22)	Partially	No	0.1
	McMahon, et al. (2016)	28 (24)	22 (16)	13.47 (2.79)	14.56 (1.61)	Yes	Yes	-0.87
	Nicholson, et al. (2020)	24 (NR)	25 (NR)	12.71 (1.52)	13.17 (1.54)	Partially	Yes	-0.66
	Williams, et al. (2018)	11 (10)	11 (8)	9.86 (1.69)	9.86 (1.00)	Yes	No	-0.79
	Wojcik, et al. (2011)[b]	16 (14)	16 (11)	11.55 (2.06)	10.95 (3.00)	Yes	No	-1.43 [a]

Category	Lead Author (year)	Participant Characteristics					Object-level	Hedges' *g*
		Sample size (males)		Mean age (std)		Matched	Matched	
		ASD	TD	ASD	TD			
Judgement-of-learning								
Adult								
	Grainger et al (2016b) Exp 1.	18 (13)	18 (11)	28.96 (10.28)	30.43 (14.59)	Yes	No	0.36
Children/adolescent								
	Grainger et al (2016b) Exp 2.	22 (19)	21 (19)	13.70 (1.45)	13.21 (1.18)	Yes	Yes	-0.22 [a]
	Wojcik, et al. (2014) Exp 1.	21 (18)	21 (17)	12.77 (2.34)	11.64 (2.49)	Yes	Yes	-1.06 [a]
	Wojcik, et al. (2014) Exp 2.	19 (NR)	19 (NR)	13.57 (2.46)	12.37 (2.56)	Yes	Yes	0.96 [a]
Feeling-of-knowing								
	Grainger et al. (2014)	18 (13)	18 (11)	28.96 (10.28)	30.43 (14.59)	Yes	Yes	-0.95
	Wojcik, et al. (2013)	18 (16)	18 (13)	12.60 (2.14)	11.83 (2.57)	Yes	Yes	-0.65 [a]

Category	Lead Author (year)	Participant Characteristics					Object-level	Hedges' *g*
		Sample size (males)		Mean age (std)			Matched	Matched
		ASD	TD	ASD	TD			
Judgement-of-performance								
	Furlano, & Kelley (2019)[b]	30 (27)	30 (27)	13.01 (1.55)	12.62 (1.64)	Yes	Yes	-0.90 [a]
	Furlano, et al. (2015)[b]	19 (18)	22 (15)	15.21 (2.04)	14.02 (1.74)	No	No	-1.39 [a]

Note. NR = not reported

Groups were considered to be matched on key characteristics if both gender and age = <.50, and FSIQ = <.50 or both VIQ and PIQ = <.50.

[a] Average effect size; [b] Effect sizes for these studies were reversed due to lower scores indicating higher meta-cognitive accuracy.

Data Extraction and Management

Consistent with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) (Moher, Liberati, Tetzlaff, Altman, & Prisma Group, 2009), key data was extracted from each study and entered into excel. These data included: study characteristics (e.g., sample size, gender & full-scale IQ), object-level and meta-level tasks, and effect size data (e.g. means and standard deviations for metacognitive accuracy). Specific data on socioeconomic status and educational attainment levels were not recorded.

Cohen's *d* effect sizes were calculated using sample size, mean and standard deviation of metacognitive accuracy for each group using the Practical Meta-Analysis Effect Size Calculator (Wilson, 2020). In cases where sample sizes, means, and/or standard deviations were unavailable, t-values were used to calculate Cohen's *d* effect sizes. Cohen's *d* was then converted into Hedges' *g*. This was to correct for any bias as a result of small sample sizes ($N < 20$). Like Cohen's *d*, Hedges' *g* is based on the standardized mean difference, and a value ≥ 0.8 can be interpreted as a large effect, ≥ 0.5 a medium effect, and ≥ 0.2 a small effect (Cohen, 1969; Hedges, 1981).

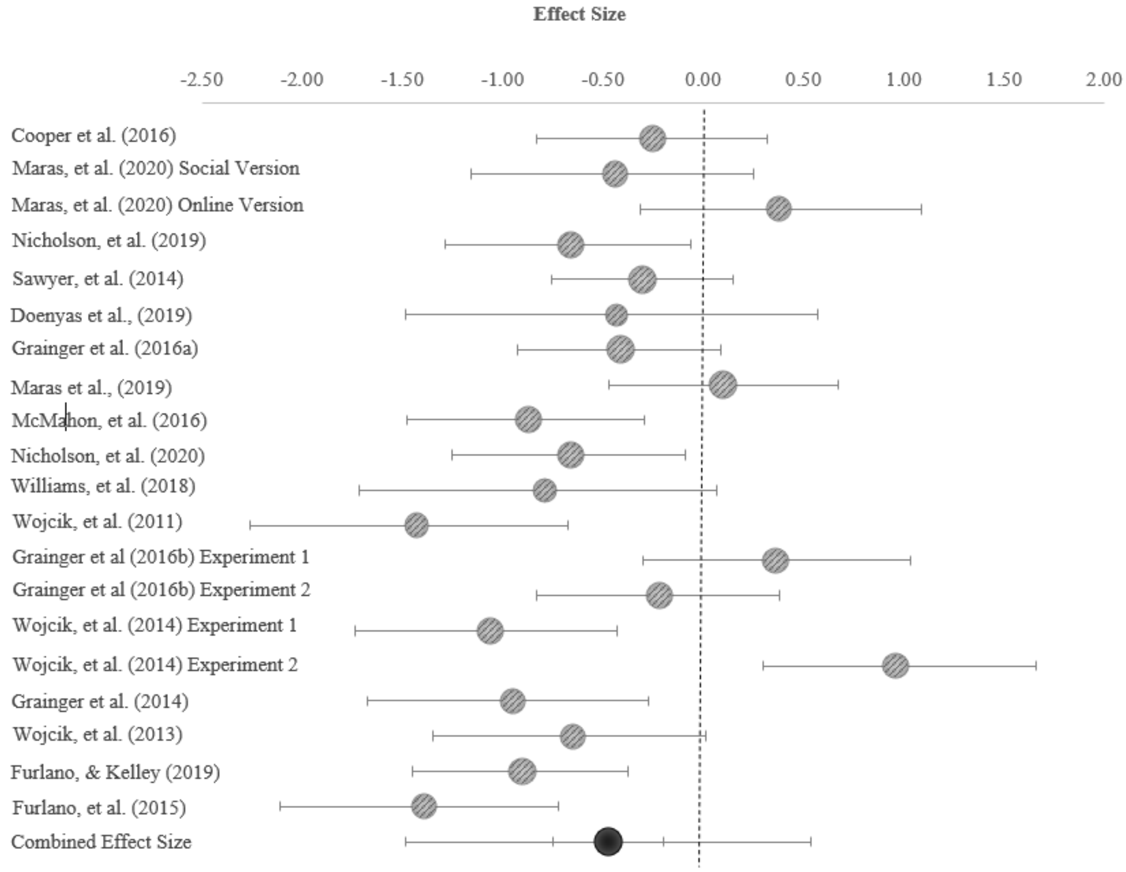
Some studies included more than one experiment/condition and therefore they also provided more than one effect size. In some cases, these effect sizes derived from the same participant group. When this was the case, an average effect size for between-group differences in metacognitive accuracy was calculated and this average effect size was then included in the meta-analysis. This is a standard procedure to manage multiple dependent effect sizes and takes into account the issues relating to using multiple dependent effect sizes within one meta-analysis (see Borenstein, Hedges, Higgins, & Rothstein, 2009). In cases where the effect sizes derived from different participant groups, these were kept separate and therefore some studies have multiple independent effect sizes included in the current meta-analysis. Overall, this approach produced 20 effect sizes that derived from experiments that fitted the inclusion criteria outlined above. Table 2 shows the key data and effect sizes for each of these experiments. Following this initial data collection, effect sizes (Hedges' *g*) and sample sizes were entered into the software package Meta-Essentials (Suurmond, van Rhee & Hak, 2017), results of which are presented below.

Results

A total of 412 autistic and 453 typically developing participants were included in the meta-analysis, and a random effects model was used. The weighted effect size for the between-group difference in meta-monitoring ability was -0.47 (SE 0.13, 95% CI -0.75 to -0.20) and statistically significant, z = -3.57, $p<.01$. This suggests a close to moderate impairment of metacognitive accuracy among the autism groups in comparison to the typically developing groups. However, the homogeneity test was significant ($Q = 61.06$, $p = <.001$), indicating that the variance across the effect sizes was greater than expected by sampling error. This suggests that there was a large range of effect sizes and so it is possible that breaking the studies down into subgroups may be more appropriate than examining them as a whole. I^2 was also large (68.88%) supporting the need for further subgroup/moderation analysis. The effect sizes and accompanying confidence intervals are presented in Figure 5. Values below zero indicate that the typically developing group performed more accurately than the autism group. The analysis was also conducted with unaveraged effect sizes (see Appendix 1), which made little difference to the results. The weighted effect size remained close to moderate and statistically significant, and the homogeneity test was still significant.

Figure 5

Forest Plot Indicating Effect Sizes and Confidence Intervals for the Main Meta-analysis



One explanation for the heterogeneity is that deficits in metacognitive accuracy are domain-specific, rather than domain-general. Evidence indicates that different meta-level tasks rely on distinct processes (Fleming, Massoni, Gajdos, & Vergnaud, 2016). Given the variety of meta-level tasks used across the studies, we performed a subgroup analysis based on meta-level task. The results showed that the weighted effect size for the between-group difference in metacognitive accuracy on judgements-of-confidence tasks was -0.45 (95% CI -0.71 to -0.20), whereas it was 0.01 (95% CI -0.84 to 0.85) for judgements-of-learning. The homogeneity test for both these subgroups were significant

(Q = 20.47, p = .04 and Q = 20.57, p = <.001, respectively). To make sense of the heterogeneity, these studies will be examined in more detail in part two of this paper. We will also examine the feeling-of-knowing and judgements-of-performance studies, of which there were too few to interpret meaningfully from the subgroup analysis, although for context the weighted effect sizes were -0.80 (95% CI -1.09 to -0.50) for feeling-of-knowing and -1.10 (95% CI -1.57 to -0.62) for judgements-of-performance, both of which were homogenous (Q = 0.39, p = .53 and Q = 1.27, p = .26, respectively).

Another reason for the heterogeneity in the initial meta-analysis could be that it combined studies of children and adults. Metacognition has been shown to have a developmental link (Weil et al., 2013), therefore, it makes sense to examine the adult and children studies independently. Out of the 20 effect sizes, 13 derived from experiments examining metacognitive accuracy in children, with the remaining seven using adult participants. Subgroup analysis showed that the weighted effect size for the between-group differences in adults was -0.27 (95% CI -0.62 to 0.08) and for children it was -0.59 (95% CI -0.93 to -0.24), indicating that the deficit in metacognitive accuracy is twice as large among children as among adults. Possible reasons for this apparent developmental difference will be examined in Part 2. The homogeneity test for both adults and children was significant (Q = 13.04, p= .04 and Q = 48.89, p=<.001, respectively). I^2 was also large for both these groups suggesting further subgroup analysis would be useful, perhaps in terms of examining it across different domains. However, due to the relatively small number of studies it would not be valid to break the analysis down any further.

Overall, these results show that there appears to be a difference in metacognitive accuracy, with most effect sizes indicating that the autism groups have poorer performance than the typically developing groups, albeit with a wide range of effect sizes. This meta-analysis makes a valuable contribution to the literature and has relevance for both theory development and clinical practice. Nevertheless, the result of a meta-analyses are only as valid and reliable as the results from the studies that comprise the analysis. Certainly, there are several issues that require consideration when interpreting case-control studies of metacognition and so Part 2 presents a critical review of the studies included in the meta-analysis.

Part 2: Critical review of Studies Methodological and Conceptual Issues in the Study of Metacognitive Accuracy in Autism

Group Matching

To draw any firm conclusions regarding differences in group performances it is important that groups are matched for key abilities/characteristics that are likely to relate to the dependent variable (i.e., metacognitive accuracy). Without matching, it is not possible to say with certainty if any between-group differences in metacognitive accuracy are the result of a true differences due to diagnostic status or just down to differences related to the extraneous variables. Failing to match groups on key characteristics makes type 1 errors more likely and should be avoided. To consider groups as equated, it has been suggested that Cohen's *d* should be <0.50 and where applicable Phi should be <0.3 (McCartney & Burchinal, 2006; Mervis & Klein-Tasman, 2004). Evidence shows that chronological age (Palmer, David & Fleming, 2014), gender (Weil et al., 2013), and IQ (Ohtani & Hisasaka, 2018) all relate to metacognitive ability. Therefore, it is important that groups are matched on these aspects when examining metacognition among individuals diagnosed with autism in comparison to typically developing individuals.

Examination of the studies included in the meta-analysis reveals that 13 out of the 20 effect sizes derived from experiments that matched groups on chronological age, gender and IQ (See Tables 1 and 2). Eight of the five studies indicated that the autism group had diminished metacognitive accuracy, with a moderate to large effect (-1.43 to -0.65). Out of the remaining five, four indicated little difference in accuracy with effect sizes ranging from -0.25 to 0.38 (Cooper, Plaisted-Grant, Baron-Cohen, & Simons, 2016; Grainger, Williams, & Lind, 2016b, experiments 1 and 2; Maras, Norris, & Brewer, 2020, online condition), and one indicated that the autism group performed better than the typically developing group with a large effect (0.96; Wojcik, Waterman, Lestié, Moulin, & Souchay, 2014, Experiment 2). Of the four that indicated little between-group difference, two used judgements-of-learning as a measure of metacognition (Grainger et al, 2016b, experiments 1 and 2). It is therefore possible that this type of meta-level judgment is undiminished among autistic individuals. The other two experiments examined judgements-of-confidence in adults (Cooper et al., 2016; Maras, et al., 2020, online condition), which may indicate that this type of meta-level judgment is undiminished among autistic adults. Both these issues will be explored in more detail below.

Object-level task Performance

In addition to ensuring that participants are matched on background characteristics it is also important that groups are matched on their object-level performance (Schwartz & Metcalf, 1994). This is because the object-level performance is involved in the computation of metacognitive accuracy. Therefore, when differences in object-level performance is taken into account it can eliminate group differences in meta-level performance (Connor, Dunlosky & Hertzog, 1997). Gamma, for example, is influenced by object-level performance, that is a person may appear to have better metacognitive accuracy on one task compared to another simply because they found one task easier (i.e., the easier the task is the easier it is to spots one's mistake).

Out of the 20 effect sizes, 15 derived from participants matched on object-level performance (Cohen's *d* <0.50). Of these 15, 10 showed a difference in metacognitive accuracy with an effect size ≥-0.41, the majority of these also matched for the key characteristics discussed above (see Tables 1 and 2). Of the remaining five, four showed very little difference in metacognitive accuracy and one indicated that the autism group performed better that the typically developing group. Overall, this suggests that even when we exclude studies that fail to match on object-level performance there continues to be a deficit in metacognitive accuracy among autistic participants in comparison to their typically developing counterparts.

Type of meta-level task

Another factor that requires consideration is the type of meta-level task used as a measure of metacognitive accuracy. Examining the outcomes from different meta-level tasks allows us to get a better understanding of the metacognitive profile of autistic individuals, and it allows us to see if any deficit in metacognitive accuracy is domain-general or specific. Evidence indicates that different meta-level task rely on distinct processes (Fleming, Massoni, Gajdos, & Vergnaud, 2016), and therefore it is possible that autistic individuals may be impaired in some meta-level tasks but not others. Thus, we conducted a subgroup analysis splitting the effect sizes up based on meta-level task.

The majority of the studies included in the main meta-analysis examined judgements-of-confidence. This analysis revealed that even when excluding other meta-level tasks there remained a close to moderate difference in accuracy as measured by judgements-of-confidence, with a wide range of effect sizes. Looking at the judgements-of-confidence studies individually, five examined judgements-of-confidence in adults, with all except one (Nicholson, Williams, Grainger, Lind, & Carruthers, 2019) showing little difference in between-group metacognitive accuracy. The remaining seven effect sizes came from children/adolescent studies, of these, six showed an effect size of $\geq$ -0.40 for the between-group difference in judgements-of-confidence accuracy. This suggests that whilst judgements-of-confidence accuracy appears to be diminished among autistic children this difficulty may have resolved by adulthood.

Turning our attention to the remaining studies, we can see that autistic children also appear to struggle with making global judgements about their cognitive-level performance. The two studies that examined judgements-of-performance both found large effect sizes among children (-0.90 and -1.39). To date no study has examined judgements-of-performance in autistic adults. Feeling-of-knowing, however, appears to be diminished among both autistic adults and autistic children. Thus far, the two studies examining feeling-of-knowing (one in adults and one in children) have both found metacognitive accuracy to be impaired with a moderate-to-large effect (-0.95, -0.65 respectively). Judgements-of-learning accuracy, however, appears to be undiminished in autism. Out of the four experiments that examined judgements-of-learning, none reported a significant difference between groups.

Overall, this suggests that autistic individuals do appear to have diminished metacognitive accuracy across a variety of meta-level judgements including confidence judgements, feeling-of-knowing judgements, and judgements-of-performance. This contrasts with judgements-of-learning accuracy, for which there is no reason to suspect any diminution in accuracy. There does, however, appear to be some distinction between autistic adult's and autistic children's metacognitive accuracy, and therefore this will be explored in more detail below.

Child versus Adult Metacognitive Performance

Further inspection of the studies included in the meta-analysis revealed that 13 out of the 20 effect sizes derived from experiments involving children and/or adolescents, with the remaining seven using adult participants. As can be seen from the subgroup analysis, when these are broken down it seems there is a small (-0.27) difference in metacognitive accuracy between autistic and typically developing adults, but a moderate (-0.59) difference between autistic and typically developing children.

Focusing on the effect sizes from the adult studies, five out of seven showed a small-to-moderate difference, three indicating poorer performance among the autistic groups (-0.44, -.30, -.25) and two indicting poorer performance in the typically developing group (.38, .36). Four of which examined judgements-of-confidence and one examined judgements-of-learning. The remaining two effect sizes were moderate to large (-0.66 and -0.95) and in both cases the typically developing group were more accurate than the autism group, one of which examined feeling-of-knowing (Grainger, Williams, & Lind., 2014) and one examined judgements-of-confidence (Nicholson et al., 2019).

Turing our attention to the effect sizes that derived from the child studies, we can see that whilst they were variable in size, they consistently favoured the comparison groups over autism groups. Ten out of the 13 effect sizes indicated the autism groups were less accurate, with effect sizes ranging from -1.43 to -0.41. Of the remaining three, two showed little difference in metacognitive accuracy (Grainger et al., 2016b, Experiment 2; Maras et al., 2019) and one showed that autism group performed better than the typically developing group, with a large effect size (0.96, Wojcik et al., 2014, Experiment 2).

The subgroup analysis and examination of these studies more closely suggest it may be sensible to examine adults and children separately when drawing any conclusions regarding between-group differences in metacognitive ability. Further studies of metacognition in autistic adults would be useful to confirm the between-group difference across a variety of object-level and meta-level tasks. It is possible that the relatively clear diminution of (most aspects of) metacognition among autistic children may not persist into adulthood. One possibility is that early metacognitive impairments resolve over development in autism. An alternative possibility is that autistic adults perform *relatively* well on metacognitive tasks through the use of compensatory strategies and/or learning/development, despite atypical underlying metacognitive competence (possible alternative strategies are discussion in more detail in chapter three). Compensation is widely believed to occur among autistic people and so it is plausible that differences between autistic and neurotypical people diminish over time because of compensation. Future studies could investigate these possibilities by exploring fine-grained patterns of performance on metacognitive tasks (as well as associations with other aspects of cognition/real-world functioning), rather than focusing only on level of metacognitive accuracy per se. If compensation underpins the relatively undiminished metacognitive accuracy observed among autistic adults, then patterns of performance (at a trial-by-trial level, for example) should still be less stable/differ from those seen among neurotypical individuals. Likewise, established links between metacognition and aspects of cognition and/or behaviour (e.g., between metacognition and general intelligence or educational achievement), should be significantly weaker among autistic than neurotypical adults.

Table 3

Details of Group Matching

Category	Lead Author (year)	Participant Characteristics					Object-Level Performance
		Gender	Age	VIQ	PIQ	FSIQ	
Judgement-of-confidence Adult							
	Cooper et al. (2016)	Yes	Yes	Yes	Yes	NR	Yes
	Maras et al. (2020) Social version	No	Yes	No	Yes	Yes	Yes
	Maras et al. (2020) Online version	Yes	Yes	Yes	No	Yes	Yes
	Sawyer et al. (2014)	Yes	No	Yes	No	Yes	Yes
Judgement-of-confidence Children/adolescent							
	Doenyas et al., (2019)	Yes	Yes	NR	NR	NR	Yes
	Grainger et al. (2016a)	NR	Yes	Yes	Yes	Yes	Yes
	Maras et al. (2019)	Yes	Yes	NR	NR	NR	No
	McMahon et al. (2016)	Yes	Yes	Yes	Yes	NR	Yes

Category	Lead Author (year)	Participant Characteristics					Object-Level Performance
		Gender	Age	VIQ	PIQ	FSIQ	
Judgement-of-confidence							
Children/adolescent							
	Nicholson et al. (2020)	NR	Yes	Yes	Yes	NR	Yes
	Williams et al. (2018)	Yes	Yes	Yes	Yes	NR	No
	Wojcik et al. (2011)	Yes	Yes	NR	NR	Yes	No
Judgement-of-learning							
Adult							
	Grainger et al (2016b) Exp 1.	Yes	Yes	Yes	Yes	Yes	No
Children/adolescent							
	Grainger et al (2016b) Exp 2.	Yes	Yes	Yes	Yes	Yes	Yes
	Wojcik et al. (2014) Exp 1.	Yes	Yes	No	Yes	Yes	Yes
	Wojcik et al. (2014) Exp 2.	Yes	Yes	No	Yes	Yes	Yes
Feeling-of-knowing							
	Grainger et al. (2014)	Yes	Yes	Yes	Yes	Yes	Yes
	Wojcik et al. (2013)	Yes	Yes	No	Yes	Yes	Yes

Category	Lead Author (year)	Participant Characteristics					Object-Level Performance
		Gender	Age	VIQ	PIQ	FSIQ	
Judgement-of-performance							
	Furlano & Kelley (2019)	Yes	Yes	Yes	Yes	Yes	Yes
	Furlano et al. (2015)	No	No	Yes	No	No	No

Note. NR = not reported

Groups were considered to be matched if Cohen's *d* = <0.50 or Phi = <0.3

General Discussion

This meta-analysis showed that there is a moderate, albeit heterogeneous, diminution of metacognitive accuracy among individuals diagnosed with autism. This was further supported by the critical review that revealed that even when key characteristics and object-level task performance was taken into account the majority of studies showed diminished metacognitive accuracy among the autistic participants. Nevertheless, the subsequent subgroup analysis and critical review showed that the level of metacognitive accuracy may vary as a result of the meta-level task being employed. For example, there is no reason to suspect that judgements-of-learning are diminished among autistic adults or children but there is clear evidence for difficulties in feeling-of-knowing judgements among both autistic adults and children.

The variation in accuracy across meta-level tasks highlights the need to explore the landscape of strengths and weaknesses in metacognitive accuracy among autistic individuals. To date, whilst some studies have varied the object-level task within the same participant group no study has yet examined different meta-level tasks within the same participant group using the same object-level task. Therefore, research that examines the various types of meta-level tasks within the same participant group would help expand our understanding of metacognition within autistic individuals. This may then inform the development of any future targeted intervention or training programmes.

The subgroup analysis and critical review of the individual meta-level tasks also highlighted the distinction between studies that involve adults and children. For example, when examining the studies that employed judgements-of-confidence, it appears that whilst autistic children may be impaired, autistic adults may in fact have intact metacognitive accuracy. Further subgroup analysis of all the meta-level tasks combined revealed that when the adult and child studies were examined separately, the diminution in metacognitive accuracy among children was moderate, but the difference between autistic and non-autistic adults was small. This was further supported by the critical review that showed that most of the child studies indicate diminished metacognitive accuracy among the autistic participants in comparison to typically developing children. This was in contrast to the adult studies where the majority of them showed little difference in metacognitive performance. Overall, this suggest that whilst autistic children may have metacognitive difficulties in some meta-level tasks, these difficulties may resolve by adulthood.

Establishing if the reduction in disparity is due to developmental delay or compensation is important because it may inform what strategies can successfully be employed to improve metacognitive accuracy among individuals who have difficulties with such tasks. Shedding light on successful strategies can help inform the development of effective and targeted intervention or training programmes that aim to improve metacognitive accuracy. This is important given that metacognition pervades daily life from the basic decisions we have to make every day to the level of our academic success and subsequent impact this has on life chances (Dunlosky & Metcalfe, 2009; Hartwig & Dunlosky, 2012; Veenman, Van Hout-Wolters, Afflerbach, 2006).

Overall, this meta-analysis and review highlights the complexities of examining metacognitive accuracy among autistic individuals. It shows that researchers and clinicians need to pay close attention to the specific areas of metacognition being examined as well as the characteristics of individuals they are examining. It also opens avenues for future research in respect to the developmental trajectory of metacognitive accuracy, the profile of strengths and weaknesses, and the effective strategies used to make accurate metacognitive judgments, particularly among autistic individuals. All of this can inform how we understand metacognition from both a theoretical and clinical perspective, which is highly important given the impact that metacognitive accuracy can have on daily life.

Chapter Three: Putting Your Money Where Your Mouth is: Examining Metacognition in ASD using Post-Decision Wagering

Abstract

Background: Following on from the previous chapter that highlighted the variation in metacognitive accuracy across different meta-level tasks, the study presented in the current chapter examines metacognition (accuracy and speed) in relation to autism and mindreading using post-decision wagering (PDW). PDW is considered to be the behavioural equivalent to judgements-of-confidence, and to date no study has examined PDW among autistic individuals. Employing behavioural measures to examine metacognitive accuracy is crucial in extending our understanding of metacognition among autistic individuals considering that many of our day-to-day meta-level judgements do not necessarily involve explicit verbal responses. ***Method*:** In Experiment 1, 39 students completed a perceptual discrimination task before making their meta-level judgement. In Experiment 2, 21 autistic adults and 20 age- and IQ-matched typically developing adults also made meta-level judgements on a similar perceptual discrimination task. ***Results:*** Results from the student sample showed negative associations between autism traits and metacognitive accuracy, and metacognitive reaction times and mindreading. These findings were replicated in a general population sample, providing evidence of a reliable association between metacognition, mindreading and autism traits. Experiment 2 showed that autistic adults have equivalent levels of metacognitive accuracy to typically developing adults, albeit with an overall increase in meta-level processing time.

Autism spectrum disorder (ASD) is a developmental condition characterised by the early onset of behavioural difficulties in social communication, and restricted/repetitive behavior and interests (American Psychiatric Association, 2013). It is widely reported that, at the cognitive level, mindreading (the ability to attribute mental states to others; also known as theory of mind or mentalising) is diminished among individuals diagnosed with ASD (Brunsdon & Happé, 2014). However, there is relatively little research focusing on metacognition (the ability to attribute mental states to oneself) among individuals with autism (Carruthers, 2009).

This relative lack of research into metacognition in ASD is surprising for several reasons. First, from a practical/clinical perspective, metacognition plays a vital role in everyday self-regulation (especially in education/work settings; Hacker et al. 2008; Nelson and Narens 1990; Schunk 2008), allowing one to control their thoughts and behaviour efficiently. For example, knowing that we do not know something should lead us to communicate our uncertainty (Bahrami et al. 2010), or seek out new information (Metcalfe and Finn 2008; Metcalfe 2009). These skills are important when it comes to real world situations, such as those faced in education or work. In these situations, uncertainty may lead one to revise more for an exam or to ask their supervisor for more guidance so that they can perform their job successfully. This is particularly relevant for understanding ASD, given that people with this disorder tend to have difficulties with self-regulation (Gomez & Baird 2005; Jahromi et al., 2013), under-achieve in education relative to what would be predicted based on general intelligence (Ohtani & Hisasaka 2018), and struggle to maintain long-term employment (Hendricks 2010; Shattuck et al. 2007).

Second, from a theoretical perspective, there remains a debate concerning the underlying cognitive processes involved in mindreading and metacognition. On the one hand, it has been proposed that mindreading and metacognition rely on the same neurocognitive mechanism, and therefore metacognition (as well as mindreading) should be impaired in individuals with autism (Carruthers, 2011). However, others have argued that mindreading and metacognition rely on distinct processes (Nichols & Stich, 2003). Given that mindreading is known to be diminished among individuals with autism,[2] the study of metacognition in ASD has the potential to contribute to theory-building in this area. For example, if mindreading and metacognition rely on the same neurocognitive mechanism it would be predicted that there would be a significant relation between individuals' performance on mindreading and metacognition tasks. The argument of shared mechanisms would further be supported by evidence of diminished metacognitive ability among individuals with autism. If, however, a dissociation is found, this would oppose the argument that mindreading and metacognition rely on the same processes and support the theories that suggest distinct or additional processes are at work.

Metacognition is assessed traditionally by asking individuals to make some form of judgement about their ability/ performance. The closer the correspondence between a person's judgement of their ability and their actual ability, the better a person's metacognitive monitoring ability. Probably the most frequently used task is the classic Judgement of Confidence (JoC) task. In this task, participants make a cognitive-level (or "object-level") judgement and then rate how confident they are that they have performed accurately. The extent to which a participant's confidence rating corresponds to the actual performance indicates their metacognitive accuracy.

To date, five studies have examined JoC among children/ adolescents with autism, four of which reported diminished metacognitive accuracy (Wilkinson et al. 2010;

[2] At the time of publication reviewers requested that we use person first language, thus this chapter contrasts with the terminology used in the rest of this thesis.

McMahon et al. 2016; Williams et al. 2018; Grainger et al. 2016), and one of which reported no significant between-group differences (Wojcik et al. 2011). A further four studies have explored metacognition among adults with ASD, three finding metacognition to be undiminished (Wilkinson et al. 2010; Sawyer et al. 2014), one producing mixed results (Cooper et al. 2016), and one reporting a significant diminution of JoC accuracy among participants with ASD (Nicholson et al. 2019). From this limited number of studies, it is difficult to draw any firm conclusions relating to metacognition as measured by JoC in adults with autism. One possibility is that metacognitive deficits in childhood are resolved by adulthood. Another possibility is that methodological (or other) issues mask underlying deficits among adults with autism. Sawyer et al. for example, did not match groups for age or IQ. Furthermore, Cooper et al. suggest that their mixed results may have been due to the differences in object-level tasks rather than true metacognitive differences. Given these mixed results, further research is required to rectify the methodological issues and examine metacognition using object-level tasks where individuals with autism do not have specific deficits.

A more general issue to consider when interpreting results from studies of JoC accuracy in ASD is that such tasks rely on verbal reports of confidence. One potential difficulty with such verbal measures is that they rely on a subjective interpretation of “confidence”, which may vary across individuals in a way that is not measured in traditional JoC tasks (Sandberg et al. 2010). Although there are good reasons to employ verbal tasks as measures of metacognitive ability (see Nicholson et al. 2019), it would be beneficial to explore other types of tasks to avoid over-reliance on a single measure. An alternative measure of metacognition that has never been employed among individuals with ASD to our knowledge, is post-decision wagering (PDW). PDW is a tangible measure and has been used as an alternative to making verbal judgements-of-confidence in studies involving adults and children (Ruffman et al. 2001; Persaud et al. 2007). PDW is similar to JoC in that participants are required to make a cognitive-/object-level judgement, but instead of rating their confidence they place a bet on the accuracy of their judgement. The extent to which participants make higher wagers for correct responses and lower wagers for incorrect responses is taken to indicate their metacognitive ability. Research has also shown PDW to be as effective at measuring metacognition, providing that the impact of risk aversion is controlled for (Dienes & Seth, 2010). Risk aversion has been linked to the way that individuals wager regardless of their level of confidence. For example, individuals may wager low to avoid large losses regardless of their level of confidence. To address this, we included a standard measure of risk aversion in the current study.

To date there are no published studies using PDW to examine metacognition in relation to ASD. Given the potential benefits of PDW, the current study used a classic PDW task to investigate metacognition, and its relation to ASD and mindreading, in adult populations. In Experiment 1, we adopted an individual differences approach and explored the relations among metacognition, mindreading, and ASD traits in the general population. In Experiment 2, we employed a case–control design, to investigate metacognition and mindreading among adults with autism, as well as typically developing (TD) adults matched for age, IQ, and risk aversion. We used both metacognitive accuracy and metacognitive reaction times as measures of metacognition. Using metacognitive reaction times alongside metacognitive accuracy allows us to examine if individuals with autism/more autism traits take longer to make their metacognitive decisions. It is important to use both measures because, whilst adults with ASD may be equally as accurate, it is possible they are slower at making meta-level decisions. If there is a difference in metacognitive reaction times (independent of "object-level" reaction times) then it is possible that individuals with autism are using more effort and/or using a different process to reach levels of accuracy equal to TD individuals (Williams et al. 2009; Frith 2013; Bowler, 1992). Based on previous research and in line with the one mechanism account, we predicted that metacognitive accuracy and metacognitive reaction times (i.e., average time taken to make a PDW, as an indicator of uncertainty) would be associated significantly with both number of ASD traits (higher ASD traits = lower accuracy and slower reaction times) and mindreading ability (higher mindreading = higher accuracy and faster reaction times), independent of perceptual/object-level accuracy and reaction times.

Experiment 1: Method

Participants

Thirty-nine students (30 females) from the University of Kent took part in the experiment. Participants had a mean age of 19.10 years (SD 0.85; range = 18–21). Participants received course credits in partial fulfilment of their degree. Both experiments in the current article were ethically approved by the University of Kent's Psychology Research Ethics Committee (201715120681034775) and informed consent was obtained prior to commencing the tasks. All participants were debriefed following each session.

Materials, procedure and scoring

Wagering task

This task was implemented using PsychoPy (Peirce 2007). There were two phases to the task, the Perceptual Judgement Phase and the Wagering Phase (see Fig. 1).

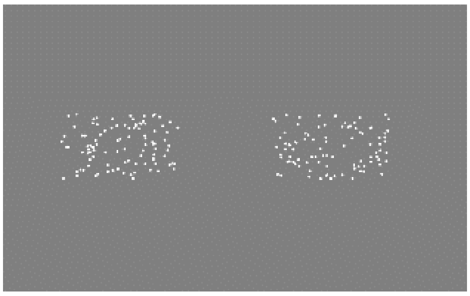
Perceptual judgement (object-level) phase. During this phase, participants were shown a series of images made up of dots (50 trials) on a computer screen. Participants were presented with two images on each trial and asked to identify which image had the most dots by clicking on the image using the mouse. The difficulty of the perceptual discrimination varied randomly across trials. Trial difficulty was operationalised in terms of the relative difference in the number of dots present in each of the two images. For example, a trial on which image A had 95 dots and image B had 125 dots (a proportional difference of .24) would be easier than a trial on which image A had 114 dots and image B had 120 dots (a proportional difference of .05). On each trial, participants had four seconds to make their judgement. If they had not made their judgement after four seconds, the programme moved on to the next trial and the trial was counted as a "miss". The proportion of trials on which a correct perceptual discrimination was made was used as one dependent variable. The average time it took participants to make their judgment was used as the second dependent variable. The quicker the discrimination response, the easier participants found it to make their judgement.

Wagering Phase. On each trial, after making their perceptual judgement, participants were asked to place a wager on their answer. Participants indicated how many counters they wished to bet by using a five-point scale. Participants were informed at the beginning of the task that if they correctly identified the image with the most dots then (a) they won back the counters they wagered plus one for every counter they wagered and (b) got to keep the counters that they did not bet. So, for example, if a participant bet three counters and their answer was correct they received the three counters they bet, plus three more and got to keep the two they had left over (thus, eight counters in total). If the participant bet three counters and their answer was incorrect, they lost the counters they wagered but got to keep the counters they had left over (i.e., if they bet three of the five counters they received two counters in total). Participants were not told how much they had won until all trials were complete. Participants could win up to 500 counters. Participants were informed that the top three people with the most counters at the end of the study would receive a prize (1st = prize worth £20, 2nd = prize worth £10 and 3rd = prize worth £5). Prior to commencing the trials participants completed 10 practice trials.

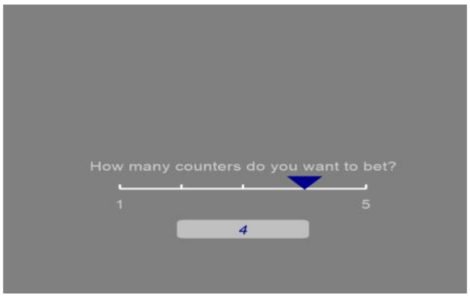
"Meta-level" performance was indexed in each participant by calculating a gamma correlation (Kruskal and Goodman 1954) between perceptual discrimination accuracy (correct/incorrect) and number of counters wagered, providing a measure of metacognitive accuracy. This measure has been recommended by Nelson (1984), and Nelson et al. (2004) and has been extensively used in research on metacognitive monitoring processes (e.g. Grainger et al. 2016; Sawyer et al. 2014; Williams et al. 2018). Use of gamma in the current study also serves to facilitate comparisons with other studies of metacognition in ASD, which have almost exclusively employed gamma as the main dependent variable. Metacognitive accuracy ranges from − 1 to + 1 with scores of 0 indicating chance level accuracy, and large positive scores indicating good metacognitive accuracy. "Meta-level" performance was also indexed by the average time it took for participants to place their bet across trials. The quicker the wagering response, the easier participants found it to make their judgement. One participant was excluded from all analysis because there was no variation in their wagers across trials and so a gamma score could not be calculated (leaving n = 39).

Figure 6

Example Trial in the Wagering Task



Perceptual Judgement Phase - Participants had 4 seconds to select the image with the most dots.



Wagering Phase - Participants took as long as they needed to place a wager on their perceptual judgement being correct.

Autism-spectrum Quotient

The Autism-spectrum Quotient (AQ) is a widely-used and well-validated self-report measure of ASD traits. It is considered to be a reliable measure of ASD traits in both clinical and subclinical populations (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The AQ presents participants with 50 individual statements (e.g., “I find social situations easy”) and participants were asked to decide the extent to which they agreed with each statement by responding on a 4-point Likert scale, ranging from “definitely agree” to “definitely disagree”. Higher scores indicate more ASD traits, with a maximum possible score of 50.

Reading the Mind in the Eyes Task

The Reading the Mind in the Eyes task (RMIE) is a widely-used measure of mindreading among intellectually able adults, including those with ASD (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). The task involves looking at photographs of eyes and deciding what the person in the picture is feeling. Participants were presented with 36 eye stimuli and were required to select an emotion that best described what the person in the picture may be feeling out of four possible emotions. Scores ranged from 0 to 36 with higher scores indicating better mindreading abilities.

It should be noted that the RMIE has been characterized reasonably as a kind of empathy/emotion recognition task, rather than a mindreading task specifically (see Oakley et al. 2016; but also see Nicholson et al. 2018). Yet, the task requires participants to select the most appropriate mental-state descriptor to explain the expression of a target agent, which appears to be a prima facie example of mindreading. The task has been employed in over 250 studies, and shows good test–retest reliability (e.g., Fernández-Abascal et al. 2013), clearly distinguishes groups of participants with and without ASD (e.g., Wilson et al. 2014), is associated with the number of ASD traits shown by individuals in large population studies (e.g., Baron-Cohen et al. 2001b), and is correlated with other measures of mindreading even after the influence of IQ is controlled statistically (e.g., Jones et al. 2018). Nonetheless, we appreciate the alternative views of the basis of the RMIE task and also that mindreading is a multi-faceted ability that may not be tapped by any single task. Therefore, we included additional measures of mindreading in both experiments 1 and 2.

Animations Task

The Animations task has been widely used to assess mindreading abilities in both the general population and those diagnosed with ASD (Abell et al. 2000). During this task, participants were required to watch four short video clips of two triangles moving around. The clips were presented on a computer screen and, after watching each clip, participants were asked to describe what they thought was happening in the video. Participants were allowed to watch each clip twice and responses were recorded using an audio recorder and later transcribed. Accurate responses required participants to attribute mental states, such as desire and intention, to the two triangles. Scores ranged from 0 to 2 for each clip, with higher scores indicating better mindreading abilities. Participants completed one practice trial prior to commencing the test trials. Inter-rater reliability across all clips was excellent according to Cicchetti's (1994) criteria (intra-class correlation = .89).

Risk Aversion tasks

Lottery questions. Participants were asked two lottery questions that were as follows (Dienes and Seth 2010):

- If there was a lottery for a £10 prize, which will be given to one of the 10 ticket holders, how much would you pay for a ticket?

- If the prize were £100, which will be given to one of the 10 ticket holders, how much would you pay for a ticket?

The smaller the amount an individual is willing to pay the lesser the individual's propensity for risk, with an optimal score of 11 indicating no risk aversion. The lottery score for our sample was not significantly different from 11, indicating that our sample was not risk averse, $t(37) = .04$, $p = .972$.

Balloon Analogue Risk Task. The Balloon Analogue Risk Task (BART) is a computer-based task designed to measure risk propensity (Lejuez et al., 2002). In this task participants were required to inflate a computer-simulated balloon by pressing the space bar. In the current study, participants earned virtual money with each pump, which was later converted into points (£1 = 1 point) and added on to their score on the wagering task. The amount earned in each trial was displayed on the screen with the total amount earned being presented throughout. When the balloon was pumped up too much, resulting in it exploding, participants did not gain anything for that trial. Participants were able to cease pumping the balloon at any point and bank the gains earned for that trial adding it to the total earnings. There were 20 trials in total. The smaller the average score for unexploded balloons the lower the individual's propensity for risk. Descriptive statistics for each of the background and risk measures are presented in Table 4.

Table 4

Experiment 1: Means and Standard Deviations for Background and Risk Measures

Variable	Mean	SD	Range
Autism Quotient	16.31	6.22	2-27
Animation	6.34	1.44	2-8
Reading the Mind in the Eyes	25.62	5.11	16-34
Balloon Analogue Risk Task	21.85	10.71	5-39
Lottery	11.05	9.18	0-55

Statistical Analysis

Reported significance values are for two-tailed tests. However, when results are predicted a priori on the basis of a solid theoretical foundation and/or previous empirical findings, it is arguably not only legitimate to use one-tailed tests, but also sensible to do so (see Cho and Abe 2013). In the current study, predictions were entirely in keeping with those made in our previous work on this topic and with published findings. Therefore, in instances where explicitly predicted results were non-significant when reported using two-tailed tests, but significant (or very close to being significant) when used one-tailed tests, we report the results from both[3]. Where t-tests were used, we report Cohen's *d* values as measures of effect size (≥ .0.20 = small effect, ≥ 0.50 = moderate effect; ≥ 0.80 = large effect; Cohen 1969). Where ANOVAs were used, we report partial eta squared (η_p^2) values as measures of effect size (≥ .01 = small effect, ≥ .06 = moderate effect, ≥ .14 = large effect; Cohen 1969).

Experiment 1: Results

Descriptive statistics for the wagering task are presented in Table 5. The gamma score on the wagering task was significantly different from zero, $t(38) = 5.50$, $p < .001$, indicating that participants were significantly above chance in their wagering accuracy, placing higher bets for correct answers than for incorrect answers. Table 6 shows non-parametric correlations among the key variables.

As predicted, AQ score was significantly negatively associated with gamma. Partial correlation analysis showed that this association remained significant even after controlling for proportion of correct object-level discriminations, $r(36) = -.34$, $p = .04$. In contrast to what was predicted, AQ score was not significantly related to wagering reaction times (RT; seconds). In terms of mindreading, RMIE was significantly negatively associated with wagering RT. This correlation remained significant even after controlling for object-level RT, $r(36) = -.41$, $p = .01$[4].

---

[3] Note that this was requested by reviewers at time of publication.

[4] Due to the unequal number of males and females within this sample, analysis was conducted to examine if there was any significant difference between males and females on key variables. This analysis showed that there was a significant difference for average wager ($t(37) = 4.15$, $p < .001$, $d = 0.41$), with males ($M = 3.5$; $SD = .51$) placing higher wagers than females ($M = 2.6$; $SD = .62$), and proportion correct ($t(37) = -$

Table 5

Experiment 1: Means and Standard Deviations for the Wagering Task

Variable	Mean	SD	Range
Object-level proportion correct	.66	.07	.56–.80
Missed trials	1.10	1.47	0–5
Object-level reaction time (s)	1.87	0.39	1.20–2.64
Counters wagered	2.78	0.71	1.00–4.44
Wagering reaction time (s)	1.60	0.39	1.00–2.83
'Meta-level' Gamma	.29	.33	-.43–1

2.67, p = .0, d = 0.17), with females (M = .68; SD = .06) getting a higher proportion of answers correct compared to males (M = .61; SD = .06). Analysis also showed that there was a marginally significant difference between males and females for RMIE ($t(37) = -1.97$, p = .06, d = 0.74), with females (M = 26.47; SD = 4.89) scoring higher than males (M = 22.78; SD = 5.07). All remaining variables, including gamma, were non-significant (all ts < 1.05, all ps > .30). Analysis was also conducted to ensure that there were no significant differences in correlations between gamma scores or wagering RT and mindreading (RMIE, Animations) or Autism traits (AQ). This analysis showed that there were no significant differences between males and females for any of the correlations according to Fisher's Z test, all $zs < -1.61$, all ps > .11.

Table 6

Experiment 1: Correlations between Key Variables

	Variables	2	3	4	5	6	7	8	9	10
1.	Object-level proportion correct	-.30	-.19	-.36*	-.01	-.29	.19	.40*	.14	-.38*
2.	Counters wagered		-.04	.02	.04	.17	-.07	-.18	.15	.06
3.	Object-level reaction times			.55**	.22	-.13	-.14	.02	.13	.32
4.	Wagering reaction times				.11	.17	-.18	-.33*	.04	.30
5.	'Meta-level' Gamma					-.32*	.13	-.04	-.05	.27
6.	Autism Quotient						-.16	-.21	-.24	.20
7.	Animation							.43**	-.21	.01
8.	Reading the Mind in the Eyes								-.20	-.26
9.	Balloon Analogue Risk Task									-.16
10.	Lottery									

Note. * p<.05, **p<.01

Experiment 1: Discussion

As predicted, the results from Experiment 1 showed that there was a significant relation between metacognitive accuracy and ASD traits, indicating that the more ASD traits an individual had the less accurate they were in their metacognitive judgements. Unexpectedly, there was no significant relation between mindreading ability and metacognitive accuracy. However, as predicted, there was a significant relation between wagering RT and mindreading as measured by the RMIE task. The better the participant's mindreading ability, the quicker they made their wagering judgements, independent of object-level RT. This implies that those with better mindreading skills are able to access metacognitive processing and interpret it quicker, and thus arrive at a wagering decision with relative ease. It should be noted, however, that wagering RT was non-significantly associated with performance on the Animations task. This could be due to the relatively limited variance in scores on the Animations task (0–8, rather than 0–36 on the RMIE task) masking an underlying association. To address these issues in Experiment 2, we employed a measure of mindreading with a wider range of scores than is possible on the Animations task (the Movie for the Assessment of Social Cognition; Dziobek et al. 2006). From these results, it was predicted that the ASD participants would show significantly lower wagering accuracy and significantly longer wagering RT than TD participants in Experiment 2.

Experiment 2: Method

Participants

Twenty-one adults with a diagnosis of ASD (13 males) and 20 TD (14 males) adults took part in the current study. All of the participants in the ASD group had received a formal diagnosis of ASD in accordance with established criteria (American Psychiatric Association, 2013; World Health Organization, 1993).

Details of participant characteristics can be seen in Table 7. Full Scale (FSIQ), Verbal (VIQ) and Performance (PIQ) IQ were assessed using the Wechsler Abbreviated Scale for Intelligence-II (Wechsler 1999). All participants also completed the AQ as a measure of ASD traits and the BART as a measure of risk aversion. Thirty-nine participants also completed the Lottery questions; the remaining two (1 ASD, 1 TD) did not due to administration error. Participants in the ASD group also completed the Autism Diagnostic Observation Schedule, a semi-structured observational measure of ASD features (Lord et al. 2000). Finally, all participants completed two measures of mindreading ability, the RMIE task and the Movie for the Assessment of Social Cognition (MASC; Dziobek et al. 2006), which is described in detail below. There were no significant differences between the ASD and TD group in terms of age, propensity for risk, FSIQ, VIQ, or PIQ. There were, however, expected between group differences in number of ASD traits (in line with their diagnostic status) and mindreading ability. Informed consent was obtained prior to commencing the tasks. All participants received payment of £7.50 per hour for their time and travel expenses, and all participants were debriefed following each session.

Table 7

Experiment 2 Participant Characteristics: Means, Standard Deviations (in brackets), and Inferential Statistics

	Group				
	ASD (n=21)	TD (n=20)	*t*	*p*	*Cohen's d*
Age	36.86 (12.22)	41.95 (13.94)	1.25	.22	0.39
Full-scale IQ	105.62 (13.18)	105.65 (12.99)	0.01	.99	0.00
Range	73-122	83-132			
Performance IQ	106.14 (16.87)	105.60 (15.18)	0.09	.93	0.04
Range	65-132	76-141			
Verbal IQ	105.38 (11.45)	104.05 (11.22)	0.38	.71	0.12
Range	86-128	81-129			
Autism Quotient	33.00 (8.20)	14.25 (4.56)	8.99	<.001	2.82
Reading the Mind in the Eyes	24.95 (5.35)	27.80 (3.86)	1.95	.06	0.61
Movie for the Assessment of Social Cognition-Total	28.10 (6.58)	33.75 (5.21)	3.04	<.001	0.95
Movie for the Assessment of Social Cognition -Control	3.43 (1.29)	4.40 (1.06)	2.60	.01	0.81
Balloon Analogue Risk Task	20.17 (9.24)	25.46 (12.36)	1.56	.13	0.48
Lottery	11.89 (24.14)	5.97 (4.30)	1.01	.29	0.36

Materials, procedure and scoring

Participants completed the AQ, RMIE, BART, lottery and wagering task all of which are described above. The procedures for AQ, RMIE and lottery were the same as in Experiment 1, although the BART involved earning money instead of points in Experiment 2. Participants also completed the MASC where they watched a short film of a group of people interacting. The film was stopped at regular intervals and the participant was asked a question about what the person in the film was thinking or feeling at the moment the film was stopped. Each question was multiple choice and participants were presented with four answers to choose from. The higher the score on the MASC the better the individual's mindreading abilities. The MASC also includes six control questions that require mental flexibility and abstract reasoning without any demand on social-cognitive competencies.

The wagering task had a similar procedure and scoring method as that used in Experiment 1, with only slight differences in each phase. In the Judgement Phase, approximately half of the participants in each group completed the same perceptual discrimination task (the dots task) as participants completed in Experiment 1. However, the other half of participants in each group completed an analogous task that involved choosing on each trial which of two lines was longest (rather than which of two boxes had the most dots in). The reason for this is that some participants had already completed the dots task as part of another study in our lab. To ensure there were no systematic differences between tasks across groups, an initial two-way ANOVA was conducted. Main effects showed that there was a significant main effect of task version, $F(1,37)$ = 5.22, p = .03, η_p^2 = .12, but not group, $F(1,37)$ = .38, p = .54, η_p^2 = .01. The task main effect indicates that participants who took part in the lines version correctly discriminated a higher proportion (.71) in comparison to the dots task (.65). Crucially, the analysis confirmed that there was no significant Group × Task version interaction on the proportion of stimuli correctly discriminated, $F(1,37) = 0.11$, $p = .72$, $\eta_p^2 = .003$.

In the Wagering Phase, the only difference in procedure in Experiment 2 from that in Experiment 1 was that money was offered instead of prizes. Hence, in Experiment 2, the number of counters participants bet was equal to the number of pennies they wish to bet, 1 counter = 1p, 2 counters = 2p and so on. One participant (with ASD) was excluded from all analysis because there was no variation in the amount they wagered across trials and so a gamma score could not be calculated. This resulted in a final ASD sample of n = 21.

Experiment 2: Results

With regard to object-level performance, there were no significant differences between participants with ASD (*M* = .67, *SD* = .08) and comparison participants (*M* = .68, *SD* = .09) in the proportion of trials on which stimuli were correctly discriminated, $t(39) = -0.44$, $p = .66$, $d = 0.12$. Moreover, there was no significant difference between the ASD group (*M* = 2.09, *SD* = .44) and TD group (*M* = 2.02, *SD* = .42) in the average number of seconds to make their perceptual judgement during the object-level phase, $t(39) = 0.48$, $p = .64$, $d = 0.16$. Thus, the two groups were very similar with respect to cognitive-/object-level ability (accuracy and speed).

In the wagering phase, there was no significant difference between the ASD group (*M* = 3.03, *SD* = .96) and TD group (*M* = 3.31, *SD* = 1.08) in number of counters wagered, $t(39) = -.91$, $p = .37$, $d = 0.27$. This confirms the findings from the BART and lottery tasks (see Table 7) that participants with ASD were not inherently more risk averse than comparison participants. Unexpectedly, the mean gamma score among participants with ASD (*M* = .37, *SD* = .26) was non-significantly smaller than the gamma score among

TD participants (*M* = .44, *SD* = .29), $t(39) = -.76$, $p = .45$, $d = 0.25$. However, as expected, the mean wagering RT was significantly longer among participants with ASD (*M* = 2.09, *SD* = .43) than among TD participants (*M* = 1.83, *SD* = .36), $t(39) = 2.08$, $p = .04$, $d = 0.66$. This remained significant (and increased somewhat in magnitude) after controlling for object-level RT, $F(1,38) = 6.70$, $p = .01$, $\eta_p^2 = .15$.

Correlations

To examine the relationship between ASD traits (AQ), metacognition and mindreading (RMIE and MASC) a series of correlational analyses were conducted among each group. In the ASD group, there were no significant correlations between wagering RT or gamma, and mindreading or ASD traits, all $r_s < .28$, all *ps* > .22. However, in the TD group, results replicated closely those observed in Experiment 1.

There was a negative correlation between AQ score and gamma among TD participants, $r_s(19) = -.42$. This correlation was close to statistical significance when using a two-tailed test, $p = .07$ and statistically significant when using a one-tailed test, $p < .04$ (which is arguably legitimate, given that it was a predicted effect). In this context, it is important to note that this correlation is actually slightly stronger than the AQ score × gamma correlation observed among TD participants in Experiment 1 ($r = -.32$ in Exp. 1 vs $r = -.42$ in Exp. 2), albeit non-significantly so according to Fisher's Z test, $z = 0.39$, $p = .35$. This suggests that the failure to reach conventional levels of statistical significance (when using a two-tailed test) was the result of the lower statistical power in Experiment 2 than in Experiment 1. Likewise, after controlling for object-level performance (proportion correct), the AQ score × gamma correlation in Experiment 2 was non-significant when using a two-tailed test, $r_s(17) = -.37$, $p = .12$, but marginally significant when using a one-tailed test, $p = .06$. Again, the partial AQ score × gamma correlation in Experiment 2 was slightly stronger than the equivalent partial correlation in Experiment 1 ($r = -.34$ in Exp. 1 vs $r = -.37$ in Exp. 2). All other analyses examining the relationships between gamma scores and mindreading for the TD group were non-significant, all $r_s < .17$, all *ps* > .24.

In terms of wagering RT, among TD participants, there was a significant negative correlation between wagering RT and performance on the MASC, $r_s(18) = -.73$, $p < .001$. This remained significant when controlling for object-level RT, $r_s(17) = -.73$, $p < .001$ and proportion correctly discriminated, $r_s(17) = -.69$, $p < .001$. All other analyses examining the relationships between wagering RT, and mindreading (RMIE) and ASD traits were non-significant, all $r_s < -.14$, all *ps* > .55.

Due to the relatively small sample sizes across the two experiments we combined the student sample from Experiment 1 and the TD sample from Experiment 2 (n = 59) to increase statistical power. Post hoc analysis using G*Power 3.1 (Faul et al. 2007) revealed the statistical power for detecting a medium effect size (.3) for the combined samples was .77. The combined sample analysis revealed that the significant negative correlation between AQ score and gamma ($r_s(59) = -.32$, $p = .01$) remained significant when controlling for object-level performance (proportion correct), $r_s(56) = -.31$, $p = .01$. There remained no significant correlation between gamma and RMIE, $r_s(59) = .05$, $p = .36$. Combining the data also showed that there was a marginally significant negative relationship between wagering RT and RMIE, $r_s(59) = -.21$, $p = .06$ (which was significant of reported using a one-tailed test, $p = .03$), but the relationship between wagering RT and AQ score for the combined samples remained non-significant, $r_s(59) = .10$, $p = .22$.

Experiment 2: Discussion

The results from Experiment 2 revealed that there was no significant difference in metacognitive accuracy between the ASD group and the TD group, in contrast to what was predicted. There was, however, a significant between-group difference in meta-level reaction time. This suggests that the ASD group may be using a different process, which requires additional processing time, to reach the same level of metacognitive accuracy as the TD group. The significant association between metacognitive accuracy and autism traits found in Experiment 1 was replicated among the TD group in Experiment 2. Furthermore, the relationship between mindreading and meta-level reaction times found in Experiment 1 was replicated in the TD group (as measured by the MASC). This suggests that individuals with poorer mindreading abilities took longer to make a metacognitive decision.

General Discussion

To our knowledge, this is the first study to investigate metacognition in relation to ASD and mindreading using PDW. The key results were that ASD traits were significantly related to metacognitive accuracy (more ASD traits = lower accuracy) and mindreading ability was associated significantly with metacognitive RT (better mindreading = faster RTs). These results, which we interpret below, should lead to the prediction that adults with a full diagnosis of ASD would show impairments in both measures of metacognitive performance (accuracy and RTs). In keeping with this prediction, wagering RTs were significantly longer among ASD participants than among TD participants in Experiment 2. In both experiments, these significant associations with meta-level performance (RTs and accuracy), were independent of the influence of object-level performance, showing the associations are specific to metacognitive, rather than cognitive, processes. In other words, it was not the case that decision-making, motor co-ordination, or general speed of processing were relatively slow among ASD participants, rather that metacognitive decision-making specifically was diminished in this sample. Perhaps most important, this pattern of associations was found in independent samples of TD adults across two experiments, which provides reassurance about the reliability of results.

Contrary to our prediction, however, there was no evidence of an ASD-specific impairment in metacognitive accuracy in Experiment 2. The between-group difference in wagering accuracy was non-significant and associated with only a small effect size (d = 0.24). This is puzzling, given the reliable association between the number of ASD traits and metacognitive accuracy in the general population. Logically, if we find a relation between variables A and B in a sample of individuals with high/clinically-significant ASD traits, then this might not necessarily hold among people with lower ASD traits, or for the general population in which high AQ scores are apparent in a small proportion of individuals. However, if the A–B correlation is reliable in the general population (which it is in our study), then it should hold in diagnosed individuals who have high ASD traits by definition. There are two possible explanations for this pattern of results, as far as we can deduce.

First, it could be that Experiment 2 was underpowered and that a larger sample of participants would have yielded a significant between-group difference in metacognitive accuracy. This is possible, of course. The sample of ASD participants was not large (which is true of many studies in the field) and so the study was not sufficiently powered to detect small/modest between-group differences. Clearly, however, the sample was sufficiently powered to detect significant between-group differences in metacognitive RTs (which were moderate in size; d = 0.65) as well as a significant association between the number of ASD traits and metacognitive accuracy among TD participants. Thus, while it may be that a larger sample would have revealed a deficit in metacognitive accuracy among individuals with autism, such a deficit would not likely be as marked as the observed deficit in metacognitive RTs and, thus, not of clinical significance, potentially.

A second explanation for the current findings is that wagering accuracy is undiminished in ASD but underpinned by slower processing efficiency in this domain which increases the amount of time people with this disorder need to make accurate metacognitive judgements. While this is a possible explanation for some of the findings, it does not appear to explain the results from the correlation analyses in Experiment 2. If performance on the wagering task was underpinned by the same underlying metarepresentational/metacognitive resources in each diagnostic group, but just resources that are accessed less quickly/efficiently among ASD than comparison participants, then associations among measures should be of a similar magnitude in each group. Yet, this was not the case. Among TD participants, wagering accuracy was associated significantly negatively with number of autism traits ($r = -.42$), but this did not hold up among participants with ASD ($r = .07$). Likewise, wagering reaction time was associated significantly with performance on the MASC measure of mindreading among TD participants ($r = -.73$), but not among participants with autism ($r = -.19$). The different patterns of association among measures in each diagnostic group suggests that the underlying processing resources used to arrive at accurate wagering decisions was different in each group. Therefore, we believe that a third explanation is more plausible, namely that participants with autism were using alternative, possibly compensatory, strategies to perform well in terms of metacognitive accuracy despite limited underlying metarepresentational competence (see Livingston & Happé, 2017).

In other domains, it has been shown that individuals with autism use alternative strategies to perform well on tasks despite atypical possessing underlying conceptual competence (Bowler, 1992; Hermelin & O'Connor, 1985). This explanation fits well with evidence that adults with autism tend to rely on deliberative reasoning strategies to solve cognitive tasks, rather than relying on intuitive processes employed by TD adults (Brosnan et al. 2016). According to Dual-Process theory (Evans & Frankish, 2009), human decision-making is underpinned by two forms of reasoning. Reasoning based on heuristics (non-analytic) tend to be fast, easy and intuitive (Type 1) and reasoning based on analytic processes tend to be slower, more effortful, and deliberative (Type 2). This notion fits well with the current findings and may also explain previous findings in the literature. Whereas TD adults from the general population tend to employ type 1 reasoning when completing metacognitive monitoring tasks, adults with autism tend to employ type 2 reasoning, which results in similar levels of accuracy but after a longer period of processing. This explains how the reliable association between ASD traits and metacognitive accuracy in the general population did not hold in the ASD sample in Experiment 2; the association we observed was between number of ASD traits and type 1 reasoning about one's confidence. ASD participants in Experiment 2 were using type 2 processing and this afforded them the opportunity to make accurate judgements despite their ASD. The previous mixed findings regarding accuracy of verbal judgements-of-confidence among adults with autism could also be explained in this way. Under some circumstances, it may be that deliberative reasoning about one's mental states yields inaccurate judgements/behaviour (and thus between-group differences in studies of monitoring accuracy). However, mostly such reasoning will yield accurate judgements (albeit after longer processing) and so between-group differences will not be observed. One potentially important issue to consider is whether there is a developmental process at work also. Intellectually-able adults with ASD have already been through an education system that encourages the development of metacognitive skills, so arguably type 2 reasoning about mental states becomes ingrained as a response to training and difficulties with intuitive monitoring earlier in life. In the context of the current study, this hypothesis would lead to the prediction that children with ASD would show significantly diminished metacognitive accuracy on the wagering task. Understanding developmental processes and not just behavioural outcomes is crucial to expanding our understanding of cognitive functioning as a whole in ASD.

From a theoretical perspective, the current results are partly in keeping with the idea that metacognition and mindreading share metarepresentational processing resources. Specifically, the speed with which one can metarepresent self (wagering reaction time) was associated specifically with the ability to metarepresent others (on the MASC and RMIE). Equally, participants with ASD showed impairments in both mindreading and metacognitive processing speed, independent of general (object-level) processing speed. These findings are consistent with the ideas that mindreading and metacognition depend on the same underlying metarepresentational resources, and that these resources are diminished in ASD causing impairments in both domains (e.g., Carruthers, 2011; Williams, 2010). Contrary to expectations, however, (a) wagering accuracy was non-significantly associated with mindreading ability, and (b) participants with ASD did not show diminished wagering accuracy. We discussed possible reasons for finding (b) above. Finding (a) was surprising, because two previous studies have reported a significant association between verbal judgement-of-confidence accuracy and mindreading abilities (Nicholson et al. 2018; Williams et al. 2018), and most assume that wagering requires the same underlying conceptual resources as judgement-of-confidence tasks, but just a different response mode. Of course, one possibility is that wagering accuracy relies on different underlying conceptual resources to judgement-of-confidence accuracy, and that only the latter requires metarepresentation (hence, only a correlation between judgement-of-confidence accuracy and mindreading task performance, but not between wagering accuracy and mindreading task performance). While this is possible, it does not explain why wagering reaction times were associated with mindreading ability. The wagering task must have tapped metarepresentational processing in some way, so it does not appear to be the case that it is not metarepresentational at all. Another possibility, therefore, is that a true underlying association in the current study between wagering accuracy and mindreading was masked by the different, non-metarepresentational demands of the tasks. The fact that the mindreading tasks employed in the current study had a verbal response mode, whereas the wagering task required only behavioural responses, may have influenced results. This idea could be tested in future studies by employing verbal and non-verbal measures of mindreading and metacognition to investigate whether specific associations exist between measures that have equivalent response modes.

Overall, the current research provides evidence that adults with ASD are just as accurate as TD adults at wagering on their perceptual judgements (implying undiminished metacognitive monitoring accuracy), albeit only with an overall increase in processing time. This is important given that metacognitive accuracy can have an impact on an individual's daily functioning (Hacker et al. 2008; Nelson & Narens, 1990), from basic tasks such as crossing a road to more complex tasks within the workplace, or even the extent to which a jury will believe a witness statement (Cutler et al. 1988).

Chapter Four: Examining Metacognition in Relation to Mindreading and Autism using Judgements of Confidence

Abstract

Background: The majority of studies examining metacognition among autistic adults has employed judgements-of-confidence as a measure of metacognitive accuracy, however no study has looked at the speed at which judgements-of-confidence are made. It is important to examine speed as well as accuracy of meta-level judgements because this may give some indication as to the ease at which autistic adults are able to reach meta-level decisions. ***Method:*** In this chapter, as with the previous chapter, two experiments are presented. In Experiment 1, 22 students completed the same perceptual discrimination task as that presented in Experiment 1 of chapter three, but instead of placing a wager, they provided a confidence rating on their answer being correct. In Experiment 2, 22 autistic adults and 20, age- and IQ-matched typically developing adults completed the same perceptual discrimination task as that used in Experiment 2 of chapter three but again rated their confidence rather than placing a wager. ***Results:*** In line with the findings from chapter three, the results from Experiment 2 revealed a significant negative association between metacognitive reaction time and mindreading. This association was of a similar magnitude in Experiment 1, albeit non-significant. There was also a significant positive association between autism traits and metacognitive reaction time among the typically developing group in Experiment 2, again this is in line with the study presented in chapter three. The results from the group analysis also replicated the findings from chapter three, with the autistic group being just as accurate as the typically developing group but slower at making their meta-level decision. These findings indicate that the associations between meta-level decision speed and both autism traits and

mindreading are persistent across meta-level tasks, as is the reduced meta-level reaction time among autistic participants. This indicates that this may be a global rather than task specific issue.

The ability to attribute mental states to *oneself* (metacognition) and the ability to attribute metal states to *others* (mindreading) are often thought to rely on the same metarepresentational processes (Carruthers, 2009; Gopnick, 1993; Leslie, 1987; Perner, 1991). If this is the case, then one would expect to find an association between mindreading and metacognition when both abilities were measured together. Conversely, one would not expect to find a dissociation between mindreading and metacognition, with only one selectively impaired in a given population.

One group of people central to understanding the relationship between mindreading and metacognition are those diagnosed with autism spectrum disorder (ASD). Autism is a neurodevelopmental condition characterised by atypical social-communication, as well as restricted and repetitive behaviours (American Psychiatric Association, 2013). Evidence suggests that difficulties with mindreading underly the apparent atypical social-communication characteristics of autism (Frith et al., 1994; Tager-Flusberg, 1999). Certainly, there is a substantial amount of evidence showing that mindreading is diminished in autistic adults and children (Baron-Cohen et al. 2001; Brewer, Young & Barnett, 2017; Brunsdon & Happé, 2014; Happé, 1995; Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). Therefore, if metacognition could be shown to be intact when mindreading is substantially impaired in this group, then it would challenge the assumption that mindreading and metacognition rely on the same inferential, metarepresentational processes. Indeed, it would support the predictions made by theorists who claim that metacognition and mindreading rely on distinct cognitive mechanisms (e.g., two-system account, Nichols & Stich, 2003), or those who claim that introspective access to one's own mental states is direct and non-inferential (e.g., simulation theory, Goldman, 2006).

Aside from the theoretical debate, understanding metacognition among autistic people is crucial from a practical perspective. If autistic people have diminished metacognitive awareness, this will have implications for the type of support provided on a day-to-day basis, as well as in education and during criminal investigations. There is clear evidence that accurate metacognitive awareness plays a pivotal role in academic success and is a highly desirable criteria when it comes to employment (Tobias & Everson, 2009; Van Laar et al. 2017; Veenman et al., 2005).

Most of the research into metacognitive accuracy in autism has focused on uncertainty monitoring ability, as traditionally measured using the classic judgements-of-confidence (JOC) task. This task requires participants to complete a cognitive (object-level) task and then make their metacognitive judgment by rating how confident they are that their object-level answer is correct. The higher the correspondence between their confidence rating (meta-level judgement) and their actual object-level performance the better their metacognitive ability. The majority of this research has found metacognitive accuracy to be diminished among autistic children/adolescents, supporting the notion that mindreading and metacognition rely on the same cognitive processes (Doenyas et al., 2019; Grainger et al., 2016; McMahon, et al., 2016; Nicholson, et al., 2020; Williams, et al., 2018). In contrast, the research among autistic adults has produced mixed findings, making it difficult to draw any firm conclusions (Cooper, Plaisten-Grand & Baron-Cohen 2016; Maras, Norrise, & Brewer, 2020; Nicholson, Williams, Grainger, Lind, & Carruthers, 2019; Sawyer, Williamson, & Young, 2014). Sawyer, et al., and Maras et al, for example, found no between-group difference in metacognitive accuracy among their groups of autistic and non-autistic adults. In contrast, Nicholson et al., found that their autism group was significantly less accurate than their typically developing group. Cooper et al., observed a more mixed set of results. In their study, autistic and non-autistic adults completed two conditions of a JoC task. In one condition (perceived-imagined) participants were required to state if they had *seen* or *imagined* a word in the previous object-level task and then rate their confidence in their answer. In the second condition (self-other) they were required to state if *they* or the *experimenter* had said the word in the previous object-level task and then rate their confidence. Cooper et al found that JoC accuracy was as high among autistic as among non-autistic participants in the self-other condition, but significantly lower in the perceived-imagined condition. Thus, the level- of accuracy appeared to depend on the requirements of the object-level task. Undoubtedly further research is required into metacognitive accuracy among autistic adults in order to clarify the theoretical debate as well as the practical implications.

To our knowledge, the only other study of uncertainty monitoring in autistic adults was conducted by Carpenter., et al. (2019; see chapter three). Rather than using a classic JoC task, Carpenter et al. employed a non-verbal post-decision wagering paradigm. Post-decision wagering is considered by some to be the behavioral equivalent of judgments of confidence (Dienes & Seth, 2010; Ruffman, Garnham, Import, & Connolly, 2001; Persaud et al., 2007). Like tasks that employ judgments of confidence, participants perform an object-level task and then make their meta-level judgement, in this case place a wager on their object-level performance being accurate. The higher the correspondence between their wager (meta-level judgement) and their actual object-level performance the better their metacognitive ability (i.e., high wagers for correct answers and low wagers for incorrect answers). Carpenter et al. found no between-group difference in metacognitive accuracy, but they did find a difference in the ease with which meta-level judgements are made. To date Carpenter et al's (2019) study is the only one to have measured the speed at which participants made their metacognitive judgments alongside metacognitive accuracy. The speed at which meta-level judgments are made is considered to reflect the effort/ease of the meta-level decisions, just as when object-level decisions are made. Carpenter at al., used a post-decision wagering paradigm to measure metacognitive awareness. From this Carpenter et al., found that the autistic participants metacognitive judgements were as accurate as their typically developing counterparts only they took longer to make such judgements. This difference persisted even when the speed of their object-level decisions was controlled for, indicating that the difference was specific to metacognitive, rather than cognitive, processes. This suggests that the autistic adults were using more effort and/or using a different process to reach the same level of accuracy as the typically developing participants. This difference in speed can have significant implications for metacognitive judgements on a day-to-day basis considering that frequency at which meta-level judgements are made combined with the additional pressures/cognitive load of daily life.

To date no study has examined meta-level judgement speed in relation to judgements-of-confidence. Therefore, the current study employed a judgment of confidence task, measuring both metacognitive accuracy and reaction time as an index of metacognitive performance. In addition, an individual differences approach was used alongside a case-control approach to clarify the findings from judgements-of-confidence studies. To date only one study has used an individual differences approach in a judgements-of-confidence task (Nicholson et al., 2019). This study found a significant association between mindreading and metacognitive accuracy, with better mindreading being associated with better metacognitive accuracy. The study did not look at meta-level speed or the association between autism traits and metacognitive accuracy/speed. The only study to have looked the association between meta-level speed, autism and mindreading from an individual differences approach used post-decision wagering and found a significant association between mindreading and meta-level speed, which indicated that individuals with better mindreading made quicker meta-level judgements (Carpenter et al., 2019). The present study aimed to see if these findings translate across meta-level tasks by utilising the same object-level task but employing a different meta-level judgement (i.e., confidence ratings).

Using an individual differences approach means that it is possible to examine autism traits in relation to other key variables. Evidence shows that milder or subclinical autism traits are common and continuously distributed within the general population (Baron-Cohen et al., 2001; Constantino & Todd, 2003). One of the major advantages of the individual differences approach is that it enables research to examine key variables (in this case mindreading, metacognition, and autism traits) independently of any comorbid conditions that are often apparent in clinical samples. For example, autistic individuals often have high rates of comorbid disorders including but not exclusively attention-deficit-hyperactivity-disorder, intellectual disability, anxiety, alexithymia (Simonoff, Pickles, Charman, Chandler, Loucas & Baird, 2008; Matson & Nebel-Schwalm, 2013). Therefore, it is possible that mixed results from autism research may be related to the lack of control over such confounding variables (Chouinard et al., 2016). The presence of such comorbid conditions, for example, makes it more difficult to attribute results specifically to autism because it is possible that the comorbid condition could have some role to play in the outcome. Thus, examining these key variables among typically developing individuals creates a way to 'control' for comorbid conditions, allowing one to gain a clearer understanding of autism specific traits and their relation to mindreading and metacognition.

Employing an individual differences approach can also be used to explore qualitative differences as well as quantitative differences. If there is a particular pattern of performance on a task that is unique to autistic participants but not related to autism traits within the typically developing population then it may indicate a causative factor in autism or an area of functioning that is severely affected by autism specifically. Intellectual disability, for example, is more common in individuals with autism but is not related to autistic traits in the general population and so it is possible that there is something specific to autism that increases the risk of having learning difficulties (Landry & Chouinard, 2016).

That is not to say that case-control studies are not vital in our understanding of mindreading and metacognition, on the contrary it is essential that both approaches are used in order to gain a broader understanding of metacognition in relation to mindreading and autism. After all, there are characteristics unique to autism that are not present among those in the general population with high levels of autism trait. Furthermore, autism requires a clinically significant impairment that interferes with everyday functioning whereas those in the general population with high levels of autism traits do not have such significant difficulties (Landry & Chouinard, 2016). Therefore, it is important to take these issues into account when applying research to everyday life, hence the rational for using a case control and individual differences approaches in the present study.

Experiment 1: Method

Participants

Twenty-two students (19 females) from a university in the southeast of England took part in the experiment[5]. Participants had a mean age of 18.91 years (*SD* .87; range = 18–21). Participants received course credits in partial fulfilment of their degree. Both experiments in the current article were ethically approved by the universities research ethics committee (201715120681034775) and informed consent was obtained prior to commencing the tasks. All participants were verbally debriefed following each session.

Materials, Procedure and Scoring

This was a laboratory study implemented using PsychoPy (Peirce, 2007) and Qualtrics software. Participants completed the Autism Spectrum Quotient (*M* = 16.41, *SD* = 6.24 as a measure of autism traits, and Reading the Mind in the Eyes (*M* = 25.55, *SD* = 5.20) and the Animation task as a measure of mindreading ability (*M* = 6.10, *SD* = 1.64). Participants also completed a judgement-of-confidence task. All tasks are described in detail below.

Judgement of Confidence task

This task was implemented in a laboratory setting using PsychoPy (Peirce, 2007). There were two phases to the task, the Perceptual Judgement Phase and the Judgement of Confidence (see Fig. 1).

Perceptual Judgement (Object-Level) Phase. During this phase, participants were shown a series of images made up of dots (50 trials) on a computer screen.

[5] Note that the intention was to collect more data for this experiment, but due restrictions related to COVID-19 the laboratories were closed and thus it was not possible to continue with data collection for this study.

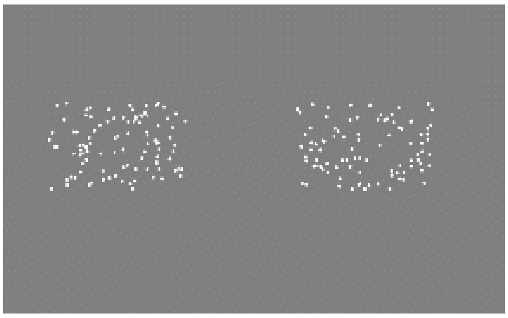
Participants were presented with two images on each trial and asked to identify which image had the most dots by clicking on the image using the mouse. The difficulty of the perceptual discrimination varied randomly across trials. Trial difficulty was operationalised in terms of the relative difference in the number of dots present in each of the two images. For example, a trial on which image A had 95 dots and image B had 125 dots (a proportional difference of .24) would be easier than a trial on which image A had 114 dots and image B had 120 dots (a proportional difference of .05). On each trial, participants had four seconds to make their judgement. If they had not made their judgement after four seconds, the programme moved on to the next trial and the trial was counted as a "miss". The proportion of trials on which a correct perceptual discrimination was made was used as one dependent variable. The average time it took participants to make their judgment was used as the second dependent variable. The quicker the discrimination response, the easier participants found it to make their judgement.

Judgement of Confidence (Meta-level) Phase. On each trial, after making their perceptual judgement, participants were asked to rate how confident they were that their answer was correct. Participants rated their confidence using a five-point scale (1 = not at all confidence, 5 = extremely confident).

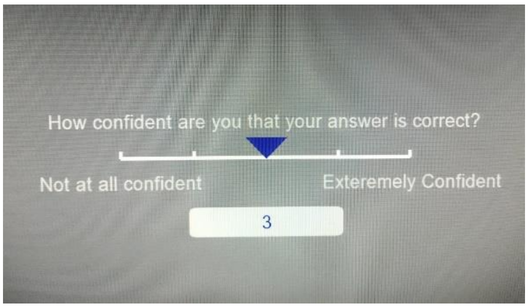
Metacognitive accuracy was indexed in each participant by calculating a gamma correlation (Kruskal & Goodman, 1954) between perceptual discrimination accuracy and number of counters wagered. This measure has been recommended by Nelson (1984), and Nelson et al. (2004) and has been extensively used in research on metacognitive monitoring processes (e.g., Grainger et al. 2016; Sawyer et al. 2014; Williams et al. 2018). Use of gamma in the current study also serves to facilitate comparisons with other studies of metacognition in ASD, which have almost exclusively employed gamma as the main dependent variable. Metacognitive accuracy ranges from − 1 to + 1 with scores of 0 indicating chance level accuracy, and large positive scores indicating good metacognitive accuracy. "Meta-level" performance was also indexed by the average time it took for participants to rate their confidence across trials. Like in Carpenter et al.'s (2019) study, the speed at which confidence judgements were made reflects the ease with which participants made meta-level judgement (i.e., the faster the judgement the easier the decision).

Figure 7

Example Trial in the Judgment of Confidence Task



Perceptual Judgement Phase: Participants had four seconds to select the image with the most dots.



Judgement of Confidence Phase: Participants took as long as they needed to rate their confidence.

Autism-spectrum Quotient

The Autism-Spectrum Quotient (AQ) is a widely-used and well-validated self-report measure of autism traits, and it is considered to be a reliable measure of autism traits in both clinical and subclinical populations (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). It is considered to be a reliable measure of ASD traits in both clinical and subclinical populations. The AQ presents participants with 50 individual statements (e.g., “I find social situations easy”) and participants were asked to decide the extent to which they agreed with each statement by responding on a 4-point Likert scale, ranging from “definitely agree” to “definitely disagree”. Higher scores indicate more ASD traits, with a maximum possible score of 50.

Reading the Mind in the Eyes Task

The Reading the Mind in the Eyes (RMIE) task is a widely used measure of mindreading ability among intellectually able adults, including those with autism (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). The adult version includes 36 photographs of the eye-region of the face. On each trial, participants are asked to pick one word from a selection of four to indicate what the person in the picture was thinking or feeling. Scores range from 0–36 with higher scores indicating better mindreading abilities. The task has been employed in over 250 studies, and shows good test-retest reliability (e.g., Fernández-Abascal, Cabello, Fernández-Berrocal, & Baron-Cohen, 2013), clearly distinguishes groups of participants with and without autism (e.g., Wilson et al., 2014), is associated with the number of autism traits shown by individuals in large population studies (e.g., Baron-Cohen et al., 2001), and is correlated with other measures of mindreading even after the influence of IQ is statistically controlled (e.g., Jones et al., 2018).

Animation task

The Animation (Abell et al. 2000) task has been widely used to assess mindreading abilities in both the general population and those diagnosed with ASD. During this task, participants were required to watch four short video clips of two triangles moving around. The clips were presented on a computer screen and, after watching each clip, participants were asked to describe what they thought was happening in the video. Participants were allowed to watch each clip twice and responses were recorded using an audio recorder and later transcribed. Accurate responses required participants to attribute mental states, such as desire and intention, to the two triangles. Scores ranged from 0 to 2 for each clip, with higher scores indicating better mindreading abilities. Participants completed one practice trial prior to commencing the test trials. Inter-rater reliability across all clips was excellent according to Cicchetti's (1994) criteria (intra-class correlation = .89).

Statistical Analysis

Where results are predicted a priori on the basis of a solid theoretical foundation and/or previous empirical findings, reported significance values are for one-tailed tests. All other significance values are two-tailed. All correlational analysis used Pearson's correlation except where the data was skewed in which case Spearman's rho was used. Where t-tests were used, we report Cohen's *d* values as measures of effect size (≥ .0.20 = small effect, ≥ 0.50 = moderate effect; ≥ 0.80 = large effect; Cohen, 1969). Where ANOVAs were used, we report partial eta squared (η_p^2) values as measures of effect size (≥ .01 = small effect, ≥ .06 = moderate effect, ≥ .14 = large effect; Cohen 1969).

Experiment 1: Results

Descriptive statistics for the wagering task are presented in Table 8. The gamma score on the judgment of confidence task was significantly different from zero, $t(21) = 5.13$, $p < .001$.This indicates that participants were significantly above chance in their overall judgement-of-confidence accuracy, rating higher confidence for correct answers than for incorrect answers.

Table 8

Experiment 1: Means and standard deviations for the Judgements of Confidence task

Variable	Mean	SD
Object-level proportion correct	.66	.11
Object-level reaction time (s)	1.88	.53
Average confidence rating	3.12	.47
Confidence rating reaction time (s)	1.53	.37
Gamma	.20	.18

Table 9 shows correlations among the key variables. In contrast to what was predicted, there was no significant correlation between AQ, and either gamma or confidence rating reaction times. In terms of mindreading, there was a moderate and marginally significant relation between Animation and confidence rating reaction times, this became significant when object-level reaction time was controlled for, $r_s(18) = -.41$, $p < .05$. There was no significant association between RMIE and confidence rating reaction times, however when object-level reaction time was controlled for this association became significant, $r_s(19) = -.37$, $p < .05$.

Table 9

Correlations between Key Variables in Experiment 1

	2	3	4	5
1. Meta-level reaction times	.16	-.12	-.36^{m}	-.20
2. Gamma		.25	-.11	-.09
3. Autism Quotient			.11	-.01
4. Animation				.50*
5. Reading the Mind in the Eyes				

Note. * p<.05, **p<.01, ***p<.001, ^{m} = Marginal (p <.06)

Experiment 1: Discussion

In contrast to what was predicted, there was no significant correlation between metacognitive accuracy and either mindreading ability or autism traits. There was, however, a marginally significant association between meta-level reaction time and mindreading, which became significant when object-level time was controlled for. This implies that those with better mindreading skills are able to access metacognitive processing and interpret it quicker, and thus make a judgement-of-confidence with relative ease. Based on these finding it was predicted that, once object-level reaction time was controlled for, there would be a significant association between meta-level reaction times and mindreading in Experiment 2. It was also predicted that there would be a between group difference in meta-level reaction times, with the autism group being slower than the typically developing group. In terms of metacognitive accuracy, although Experiment 1 did not find a significant relationship between mindreading and metacognitive accuracy, Nicholson et al., (2019), conducted a very similar experiment using a larger sample of typically developing and autistic participants and found a significant relationship between mindreading and metacognitive accuracy. Therefore, in line with this research we predict that there will be a between group difference in Experiment 2, with the autism group being less accurate than the typically developing group. Considering the debate concerning RMIE as a measure of mindreading, we also included the Movie for the Assessment of Social Cognition (Dziobek et al. 2006) as an additional measure of mindreading in Experiment 2.

Experiment 2: Method

Participants

Twenty-two autistic adults (14 males) and 20 typically developing adults (14 males) took part in the current study. All of the participants in the autism group had received a formal diagnosis of ASD in accordance with established criteria (American Psychiatric Association 2013; World Health Organization, 1993). All of the autistic participants complete the Autism Diagnostic Observation Schedule (ADOS), 16 of whom scored seven or above (M = 9.27, SD = 4.68). The ADOS is a semi-structured observational assessment of social, imagination and communication skills and a score of ≥7 is consistent with a diagnosis of autism (Lord et al., 2000).

Details of participant characteristics can be seen in Table 10. Full Scale (FSIQ), Verbal (VIQ) and Performance (PIQ) IQ were assessed using the Wechsler Abbreviated Scale for Intelligence-II (Wechsler, 1999). All participants also completed the AQ as a measure of autism traits. Finally, all participants completed three measures of mindreading ability, the RMIE task and the Animations task, as described above, and the Movie for the Assessment of Social Cognition (MASC; Dziobek et al. 2006), which is described in detail below. There were no significant differences between the autism and typically developing group in terms of age, propensity for risk, FSIQ, VIQ, or PIQ. There were, however, expected between group differences in number of autism traits (in line with their diagnostic status). There was also a significant difference in mindreading ability as measured by the MASC but not RMIE or Animations. The MASC between-group difference remained significant even when controlling for performance on the MASC control questions, $F(38) = 6.29$, $p<.05$, $\eta_p^2 = .14$. Informed consent was obtained prior to commencing the tasks. All participants received payment of £7.50 per hour for their time and travel expenses, and all participants were debriefed following each session.

Table 10

Experiment 2 Participant Characteristics: Means, Standard Deviations (in brackets), and Inferential Statistics

	Group				
	ASD (n = 22)	TD (n = 20)	*t*	*p*	*Cohen's d*
Age	36.55 (12.02)	41.95 (13.94)	1.35	.19	0.41
Full-scale IQ	106.32 (13.27)	105.65 (12.99)	.17	.87	0.05
Performance IQ	106.41 (17.51)	105.60 (15.18)	.16	.87	0.05
Verbal IQ	105.41 (11.18)	104.05 (11.22)	0.39	.70	0.12
Autism Quotient	32.86 (8.03)	14.25 (4.56)	9.11	<.001	2.85
Reading the Mind in the Eyes	25.36 (5.57)	27.80 (3.86)	1.63	.11	0.51
Animation	4.90 (2.26)	5.68 (1.57)	1.28	.21	0.40
Movie for the Assessment of Social Cognition-Total	28.32 (6.51)	33.75 (5.21)	2.98	<.01	0.92
Movie for the Assessment of Social Cognition -Control	3.50 (1.30)	4.40 (1.06)	.38	.04	0.76

Materials, procedure and scoring

Participants completed the AQ, RMIE, Animations and JoC task all of which are described above. The procedures for AQ, RMIE, and Animations were the same as in Experiment 1. Participants also completed the MASC where they watched a short film of a group of people interacting. The film was stopped at regular intervals and the participant was asked a question about what the person in the film was thinking or feeling at the moment the film was stopped. Each question was multiple choice and participants were presented with four answers to choose from. The higher the score on the MASC the better the individual's mindreading abilities. The MASC also includes six control questions that require mental flexibility and abstract reasoning without any demand on social-cognitive competencies.

The judgment of confidence task had a similar procedure and scoring method as that used in Experiment 1, with only slight differences in each phase. In the Judgement Phase, approximately half of the participants in each group completed the same perceptual discrimination task (the dots task) as participants completed in Experiment 1. However, the other half of participants in each group completed an analogous task that involved choosing on each trial which of two lines was longest (rather than which of two boxes had the most dots in). The reason for this is that some participants had already completed the dots task as part of another study in our lab. To ensure there were no systematic differences between tasks across groups, an initial two-way ANOVA was conducted. There was no main effect of task version, $F(1,8) = 1.06$, $p = .31$, $\eta_p^2 = .03$, or group, $F(1,37) = .76$, $p = .39$, $\eta_p^2 = .02$. The analysis also confirmed that there was no significant Group × Task version interaction on the proportion of stimuli correctly discriminated, $F(1,38) = 1.41$, $p = .24$, $\eta_p^2 = .04$.

Experiment 2: Results

Diagnostic Group Comparisons

With regard to object-level performance, there were no significant differences between the autism group (*M* = .68, *SD* = .09) and the typically developing group (*M* = .70, *SD* = .06) in the proportion of trials on which stimuli were correctly discriminated, *t*(40) = 0.89, *p* = .38, *d* = 0.26. Moreover, there was no significant difference between the autism group (*M* = 1.95, *SD* = .32) and typically developing group (*M* = 2.00, *SD* = .35) in the average number of seconds it took to make their perceptual judgement during the object-level phase, *t*(40) = 0.43, *p* = .67, *d* = 0.15. Thus, the two groups were very similar with respect to cognitive-/object-level ability (accuracy and speed).

In the JoC phase, there was a significant between-group difference in the average confidence rating, *t*(40) = 2.65, *p* = .01, *d* = 0.82, with the autism group (*M* = 2.03, *SD* = .38) giving higher confidence ratings than the typically developing group (*M* = 1.73, *SD* = .35). Unexpectedly, the mean gamma score among autistic participants (*M* = .39, *SD* = .26) was not significantly different to the gamma score among typically developing participants (*M* = .36, *SD* = .20), *t*(40) = 0.27, *p* = .79, *d* = 0.13. This remained non-significant when controlling for confidence ratings, *F*(1, 39) = 1.79, *p* = .20, η_p^2 = .04. However, as expected, the mean JoC RT was significantly longer among the autistic participants (*M* = 2.03, *SD* = .38) than among the typically developing participants (*M* = 1.73, *SD* = .35), *t*(40) = 2.65, *p* = .01, *d* = 0.82. This remained significant (and increased somewhat in magnitude) after controlling for object-level RT and confidence rating, *F*(1,42) = 10.53, *p* <.01, η_p^2 = .23.

Correlations

To examine the relationship between autism traits (AQ), metacognition (gamma) and mindreading (RMIE) a series of correlational analyses was conducted. All 22 ASD and 20 TD participants were included, given that the correlation analyses are essentially an individual differences approach. All results from the correlational analysis are presented in Table 11.

As can be seen in Table 11, AQ was significantly associated with meta-level reaction time. This association remained significant when controlling for object-level reaction time, $r_s(39) = .43$, $p < .01$. There was also a significant association between RMIE and meta-level reaction time. This association reduced in strength and was borderline significant when controlling for object-level reaction time, $r_s(39) = -.24$, $p = .06$. Although, it was not significantly different from the association found in Experiment 1, $z = 0.51$, $p = .30$. There was also a significant association between MASC and meta-level reaction time. This association remained significant when controlling for object-level reaction time and performance on the MASC control questions, $r_s(37) = -.42$, $p < .01$. There was no significant Animation × meta-level reaction association, and when object-level reaction time was controlled for, $r_s(37) = .13$, $p = .22$, the association was significantly smaller than the association observed in Experiment 1, $z = -2.02$, $p < .05$.

Table 11

Experiment 2 Correlations between Key Variables

Variables	2	3	4	5	6
1. Autism Quotient	-.30*	-.29*	-.00	-.04	.40**
2. Reading the Mind in the Eyes		.56***	.34**	-.02	-.28*
3. Movie for the Assessment of Social Cognition			.41**	.02	-.44**
4. Animations				-.21	.10
5. Gamma					.12
6. Meta-level reaction time					

Note. * *p*<.05, ***p*<.01, ****p*<.001,

According to Fisher's Z test there was a significant difference between the autism group and the typically developing group in the magnitude of the RMIE x meta-level reaction time correlation, z = 1.76, *p* <.05. This correlation was significant in the typically developing group, $r_s(20) = -.49$, *p* <.01, but not the autism group, $r_s(22) = -.05$, *p* = .41. The RMIE × meta-level reaction time correlation in the typically developing group remained significant when controlling for object-level reaction time, $r_s(17) = -.38$, *p* <.05. There was also a significant difference between the autism group and the typically developing group in the magnitude of the MASC × meta-level reaction time correlation, z = -2.18, *p* <.05. This correlation was significant in the typically developing group, $r_s(20) = -.69$, *p* <.001, but not the autism group, $r_s(22) = -.12$, *p* = .30. The MASC × meta-level reaction time correlation in the typically developing group remained significant when controlling for object-level reaction time, $r_s(17) = -.68$, *p* <.001. All other Zs between meta-level performance (gamma and reaction time) and AQ and mindreading (RMIE, Animations, MASC) were < 0.87 and all *p*s > .19. Whilst there was not a significant between-group difference in the association between AQ and gamma, it was noted that the association was moderately significant among the typically developing group, $r_s(20) = -.33$, $p = .08$, but not the autism group, $r_s(22) = -.06$, $p = .39$. The AQ × gamma association remained close to moderate in size but became non-significant, $r_s(17) = -.28$, p = 13, when object-level was performance was controlled for. The association for the autism group remained minimal and non-significant, $r_s(1) = -.07$, $p = .38$, when object-level was controlled for.

Experiment 2: Discussion

In contrast to what was predicted, the results from Experiment 2 revealed that there was no significant difference in metacognitive accuracy between the autism group and the typically developing group. There was however, a significant between group difference in meta-level reaction time, with the typically developing group making faster judgements than the autistic group. The results from Experiment 2 also revealed a significant association between autism, mindreading, and ease (speed) at which meta-level judgements are made. Indicting that the fewer autism traits a participant has and the better their mindreading the quicker their meta-level judgement, and therefore the ease with which such decisions are made. Further examination of the results showed that when the groups were examined separately the association between mindreading (RMIE and MASC) and meta-level reaction time was only significant in the typically developing group. There was also a marginally significant association between autism traits and gamma for the typically developing group, this remained close to moderate in size but became non-significant when controlling for object-level performance. Although as found in other studies (see chapters five and six) with larger sample sizes association of a similar/smaller size have been significant.

General Discussion

This study examined both the accuracy and ease with which confidence judgements are made in relation to autism and mindreading among two typically developing groups (student sample and general population sample) and a group of autistic adults. The key finding was the significant between-group difference in speed/ease with which meta-level judgements were made, with typically developing participants making quicker meta-level judgments than the autistic participants. This replicates Carpenter et al.'s (2019 – Chapter three) findings from their post-decision wagering study, suggesting that such differences persist across different metacognitive tasks. This conclusion is further supported by the significant association found among mindreading, and the ease with which meta-level judgements are made. In both Experiments 1 and 2, as well as in Carpenter et al's (Chapter three) study there was a significant association between meta-level reaction time and mindreading, as measured by RMIE (Exp 1: r_s = -.37; Exp 2: r_s = -.24, TD only r_s = -.38; Carpenter et al Exp 1: r_s = -.41). There was also a significant association between meta-level reaction time and mindreading as measured by the MASC among the typically developing group in the present study (r_s = -.68), replicating Carpenter et als findings (r_s = -.73).

In terms of associations between metacognitive accuracy, the current study showed no evidence for an association with mindreading or autism traits. This contrasts with what would be expected if mindreading and metacognition relies on the same cognitive mechanisms. Thus, these findings contrast with what proponents of the one-system/theory-theory would expect. This also contrasts with the findings from the previous chapter (Carpenter et al, 2019) which found a significant association between metacognitive accuracy and autism traits.

Also, in line with Carpenter et al's findings, there was no between group difference in metacognitive accuracy. This suggests that, although the autistic participants took longer to make their decisions, they were just as accurate as the typically developing participants. This is in line with predictions made by both simulation theory and the two-mechanism theory but contrasts with previous findings that have used similar object-level and meta-level tasks (Nicholson et al., 2019), and potentially challenges the one-system account. Consistent with Carpenter et al's interpretation, we propose that the autistic participants were using an alternative, possibly compensatory, strategy to achieve the same level of accuracy as the typically developing participants. Indeed, in keeping with Carpenter et al's findings, the associations between autism traits, mindreading and metacognitive reaction time in Experiment 2 were only found within the typically developing population. If it was just a case accessing resources less quickly/efficiently we would expect to find associations of the same magnitude within each group. Therefore, our results concur with Carpenter et al's conclusion that despite limited underlying metarepresentational competency, autistic adults are able to achieve the same level of metacognitive accuracy using different strategies. This conclusion is in keeping with previous research that has found that autistic people use alternative strategies to perform well on tasks despite diminished underlying competence (Bowler, 1992; Hermelin & O'Connor 1985; Livingston & Happé, 2017, Williams, 2009).

To test this hypothesis further, research is required to see if this extended duration enabled the autistic participants to reach equal levels of accuracy, and whether they would have been as accurate given the same amount of time as the typically developing participants. Future studies could, for example, restrict the time autistic participants have to make their meta-level judgements to that of a typically developing population (< 2 seconds). This would enable us to see if the autistic participants are able to make equally accurate metacognitive judgements in the same time frame.

This still leaves the question of why Nicholson et al., (2019) found a significant group difference in metacognitive accuracy but the current study did not, despite having used very similar tasks. Interestingly a key difference between the two studies is that Nicholson et al., provided feedback to their participant following each meta-level judgement. Therefore, it is possible that the typically developing sample were able to make use of this feedback more effectively than the autism group. Although previous research employing feedback has not found this to be the case and so this is unlikely to explain the conflicting findings (Maras et al, 2019; Williams et al., 2016). Another key difference is that Nicholson et al restricted the time participants had to make their meta-level judgements to 3 seconds, however, this also seem unlikely to explain the difference given that the participants in the current study only took an average of 2.03 seconds to make their judgement. Thus, the participants in Nicholson et al's study would certainly have had enough time to make their decision in the same way as they did in the present study. A more plausible explanation is the difference in pervasiveness of mindreading difficulties between the two studies. In Nicholson et al's study the autistic participants were significantly poorer at mindreading across both RMIE and Animations. In contrast, whilst there was a significant between group difference on the MASC in the current study, there was no significant between-group difference on either RMIE or Animations. Given that the one-system view predicts metacognitive abilities to be in line with mindreading abilities it is perhaps not surprising that the participants in our study were able to make relatively accurate meta-level judgements. Indeed, this is supported by the associations found between mindreading and metacognition in this and previous studies (Carpenter et al., 2019; Nicholson et al., 2019).

To test this hypothesis further, research should be conducted to examine the pervasiveness of mindreading difficulties among autistic people and the impact that this has on metacognitive accuracy. For example, future studies could recruit autistic participants with pervasive mindreading difficulties (i.e., difficulties across multiple mindreading measures) and compare them to participants with less severe mindreading difficulties (i.e., difficulty with only one particular task/area of mindreading). Certainly, this is an area worth pursuing if we are to fully understand metacognition in autism, which is vital given the impact that diminished metacognition can have on daily functioning.

Overall, these results suggest that whilst autistic adults are just as accurate in their metacognitive judgements as typically developing adults, they take longer to make their decisions. These findings replicate previous research examining meta-level speed using post-decision wagering, suggesting that this difference persists across different metacognitive tasks (Carpenter et al., 2019). Consistent with previous research the between-group difference in magnitude of associations between mindreading, autism traits and metacognition suggest that autistic people are using different, possibly compensatory, strategies to arrive at the same level of accuracy as typically developing people. This partially supports the one-system view that metacognition is diminished among autistic participants, although future research should investigate this thoroughly before any firm conclusions are drawn. Especially, given the lack of association between metacognitive accuracy and mindreading/autism traits.

The findings also raise questions over the impact that the pervasiveness that mindreading difficulties among autistic people may have on metacognitive accuracy. This is certainly another area worth investigating if we are to fully understand metacognition in autism and clarify the theoretical debate concerning the processes underlying mindreading and metacognition. Regardless of these outstanding questions, the findings from the current study should be used to inform any support provided to autistic adults, particularly when it comes to situations where they may be asked to make many meta-level judgements in quick succession.

Chapter Five: Predicting the Future and Judging the Past: Examining Prospective and Retrospective Metacognition in Relation to Mindreading in Autism

Abstract

Background: Many of our day-to day meta-level judgements are made prior to completing a cognitive-level task, yet no study into accuracy of confidence judgements in relation to autism and mindreading has examined prospective confidence judgments. This is an important omission given that evidence indicates that there may be a dissociation between prospective and retrospective meta-level judgements. ***Method*****:** In Experiment 1, 82 students participated in both a prospective and a retrospective judgements-of-confidence memory task. For Experiment 2, 27 autistic and 27 typically developing adults completed the same task as that presented in experiment 1. ***Results:*** In contrast to chapters three and four, the results from the student sample showed a significant positive association between metacognitive accuracy (prospective and retrospective) and mindreading, these finding were also replicated in Experiment 2. There was a significant negative association between prospective metacognitive accuracy and autistic traits in Experiment 1, but not 2. As with chapters three and four, there was no between group differences in metacognitive accuracy or mindreading. Although, when a between group difference was created in mindreading by using a small subsample of participants, the difference in metacognitive accuracy became moderate, although remained non-significant.

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by the early onset of social-communication difficulties and restricted/repetitive behaviours (American Psychiatric Association, 2013). The social aspects include difficulties with social reciprocity as well as maintaining and understanding relationships. The non-social aspects refer to behaviours such as stereotyped or repetitive motor movements, as well as highly restricted interests. This list is not exhaustive, and the social and non-social features of autism can manifest themselves in a variety of ways making it a highly complex and multifaceted condition.

Many theories have attempted to elucidate the cognitive underpinnings of autistic features (Baron-Cohen, Leslie & Frith, 1985; Frith & Happé, 1994; Happé & Frith, 2006; Happé & Ronald, 2008; Steel, Gorman & Flexman, 1984; Rumsey, 1985). One of the most influential of these being the theory of mind account. This theory suggests that the social difficulties characteristic of autism can be explained by a deficit in theory of mind. According to the original definition, theory of mind is the ability to attribute mental states to *oneself* and *others* in order to explain and predict behavior (Premack & Woodruff, 1978). Theory of mind requires the ability to metarepresent, or to represent mental representations (e.g., Leslie, 1987; Perner, 1991), and this ability is considered by many to be needed both in one's own case and in the case of others (Carruthers, 2009; Gopnick, 1993; Leslie, 1987; Perner 1991).

A vast body of research has consistently demonstrated a difficulty with representing mental states in others (otherwise known as "mindreading") among both autistic adults and children (e.g., Yirmiya et al., 1998), but comparatively little research has investigated the ability to represent mental states in self (otherwise known as "metacognition").

This is surprising considering that metacognition pervades almost every aspect of life, and it helps people to understand, predict and regulate their own cognitions and behaviours. Moreover, there is evidence to show that metacognition is a predictor of academic performance independent of intelligence (Hartwig & Dunlowsky, 2012; Veenman & Spaans, 2005). This is particularly important given that individuals diagnosed with autism are known to under-achieve in education relative to what would be predicted based on their general intelligence (Ohtani & Hisasaka, 2018). This becomes even more important when one takes into account the impact that academic achievement can have on employment and life chances, especially considering that autistic adults struggle to maintain long-term employment (Hendricks, 2010; Shattuck et al., 2007).

This lack of research is not only surprising from a clinical perspective but also from and theoretical perspective. The two-system account and simulation theory, for example, argue that metacognition is intact among autistic people (Goldman, 2006; Nichols & Stich, 2003). In contrast the one-system account (theory-theory) predicts that, because mindreading and metacognition rely on the same underlying processes, autistic people will have diminished metacognition as well as mindreading (Carruthers, 2009).

There are many ways to measure the accuracy of one's metacognitive awareness, but the majority of research into metacognition in autism has used retrospective judgements-of-confidence (see Chapter two). In these tasks, participants complete a cognitive-level task such answering some general knowledge questions and then rating their confidence in their chosen answer being correct. Metacognitive accuracy is then determined by the association between the confidence rating and actual performance on the cognitive-level task. The stronger the relationship the better the metacognitive accuracy (i.e., high confidence for correct answers, low confidence of incorrect answers).

Overall, this research indicates that retrospective judgements-of-confidence accuracy is diminished among autistic *children* (Doenyas et al., 2019; Grainger et al. 2016b; McMahon, et al., 2016; Nicholson, et al., 2020; Williams, et al. 2018). In contrast, the findings among autistic *adults* are much less conclusive, with two studies finding it to be intact, two finding it to be diminished, and one finding mixed results depending on the cognitive-level task (Cooper et al., 2016; Maras, et al., 2020; Nicholson, et al. 2019; Sawyer, et al., 2014; Wilkinson et al. 2010). In addition to these published studies, the study presented in chapter four of this thesis also found no significant between group differences in meta-cognitive accuracy. Thus, further investigation into judgement-of-confidence accuracy among autistic adults is required to fully understand the metacognitive competency among this group.

Interestingly, no study has examined *prospective* judgements-of-confidence in autism (i.e., confidence in a future decision). This is an important omission given that prospective and retrospective metacognitive judgments are thought by some to depend on distinct processes (Fleming, Massoni, Gajdos, & Vergnaud, 2016). Fleming et al., for example, found that retrospective judgements rely on speed and accuracy of the cognitive-level decision, whereas prospective judgements were influenced by previous levels of confidence. There is also evidence for a neural dissociation between prospective and retrospective judgements (Chua et al. 2009; Fleming & Dolan, 2012). For example, Schnyer et al. (2004) found that people with damage to the right ventromedial prefrontal cortex had intact prospective metacognitive awareness but diminished retrospective metacognitive awareness. Furthermore, research into other clinical conditions such as Alzheimer disease has found a dissociation between prospective and retrospective metacognitive awareness (see Cosentino, 2014). Considering the evidence highlighting the dissociation between prospective and retrospective metacognitive awareness it crucial that research into metacognition in autism examines both these aspects of metacognition.

To date, only four studies have examined the accuracy of prospective metacognitive judgments in autism, two using feeling-of-knowing judgments (Grainger et al., 2016; Wojcki et al., 2014) and two using judgments-of-learning (Grainger, Williams, & Lind, 2016; Wojcik, Waterman, Lestié, Moulin & Souchay, 2014; Wojcik, Waterman, Lestié, Moulin & Souchay, 2014). Tasks that employ feeling-of-knowing judgments require participants to judge the likelihood that they will know a specific piece information, that they previously failed to recall, at a future point in time (meta-level). Following the meta-level judgment participants are then tested on the ability to recall/recognise the previously unrecalled information (cognitive-level task). Judgments-of-learning tasks, on the other hand, require participants to study some information (e.g., word-pairs) and then rate how likely it is that they have learnt that piece of information. Participants are then tested on their memory for the information they were required to learn (cognitive-level task). For both these tasks, as with judgments-of-confidence tasks, the stronger the association between the cognitive-level task (actual recall/recognition) and the metacognitive judgement the more accurate the participant's metacognitive awareness.

The two studies that used feeling-of-knowing as a measure of metacognitive awareness in autism both found prospective metacognition to be diminished among autistic participants. In contrast, tasks that employed judgments-of-learning (JoL), have found metacognition to be intact. These findings to not offer a clear conclusion regarding the state of prospective metacognition in autism. This may be due to differences in meta-level judgements (JoL v FoK) or the various cognitive-level tasks, both of which can have an impact on metacognitive awareness due to the different processes and cognitive demands. Thus, further research examining prospective and retrospective metacognition using the same type of meta-level judgment and cognitive-level task is required before any firm conclusions can be drawn regarding a dissociation between *prospective* and *retrospective* awareness in autism.

Overall, the evidence presented has not provided a clear picture of metacognition in autism, with some studies showing it to be intact and some showing it to be diminished. Most notable are the inconsistent results found in retrospective judgments-of-confidence accuracy among autistic adults. In addition, the complete lack of research into the accuracy of prospective confidence judgements makes the understanding of metacognition in autism incomplete. This is particularly important considering the evidence for a dissociation in *prospective* and *retrospective* metacognitive awareness. Thus, filling these gaps in the research would expand our clinical and theoretical understanding of metacognition in autism by elucidating the extent to which metacognition is globally impaired.

In line with these recommendations, the current study used a judgement-of-confidence paradigm to examine *prospective* and *retrospective* metacognition in relation to mindreading and autism. Experiment 1 employed an individual differences approach and Experiment 2 employed a case-control approach. Autism traits span both clinical and subclinical population, with research having shown autism traits to be normally distributed across the general population (Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001; Constantino, 2011; Constantino and Todd 2000, 2003; Ronald et al., 2006). Moreover, the point at which an individual manifests enough autism traits for a formal clinical diagnosis to be given is arguably an arbitrary cut-off point, rather than a clear categorical point which divides “normality” from “abnormality”, or autistic from non-autistic (Pickles et al., 2000). These milder traits are often referred to as the broad autism phonotype (BAP). There are also various advantages to using an individual differences approach over a case-control design. The main advantage is that it significantly expands the pool of possible participants, allowing for larger sample sizes and a more readily available group of participants to draw upon (Constantino & Todd, 2003). Larger sample sizes provide greater statistical power to detect true effects (Landry & Chouinard, 2016). Indeed, small sample sizes in previous case-control studies of metacognition in autism have made it difficult to tell whether a) null findings reflect type II errors due to insufficient statistical power, or b) whether metacognitive impairments in autism (when they are observed) are related to well-established mindreading impairments in this disorder. That is not to say that case control studies are irrelevant, indeed such studies are required considering that there are characteristics unique to autism that are not found within the BAP (Landry & Chouinard, 2016). Moreover, whilst the BAP resembles autism it does not necessarily have the same underlying causes as those traits found among autistic individuals (Constantino & Todd, 2003; Landry & Chouinard, 2016; Williams et al., 2017). Therefore, to gain a clearer picture, it is legitimate and arguably advisable to use both an individual differences approach and a case control approach when examining metacognition in relation to autism and mindreading.

Experiment 1: Method

Participants

Eighty-two participants (67 females) from a university in the southeast of England took part in the experiment. Participants had a mean age of 19.5 years (*SD* 3.15; range = 18 - 46). Participants received course credits in partial fulfilment of their degree. The study was ethically approved by the University's Research Ethics Committee (201915723711285984) and informed consent was obtained prior to commencing the tasks. All participants were given a written debrief following completion of the tasks and could contact the researchers for further debrief if required.

Materials, Procedure and Scoring

This was an online study implemented using Inquisit 4 (2015). Participants were instructed to use their own laptop or desktop in a quiet space without any distractions. Participants completed the Autism Quotient (*M* = 17.49, *SD* = 6.33) and Reading the Mind in the Eyes (*M* = 25.49, *SD* = 5.68), as a measure of autism traits and mindreading ability, respectively. Participants also completed two conditions of a judgement-of-confidence task, *prospective* and *retrospective*, with order of condition counterbalanced across participants. The stimuli were 180 four-letter words with a Kucera-Francis written frequency (1967) ranging from 1-88, chosen from the MRC Psycholinguistic Database (Coltheart, 1981). For each condition, participants completed three blocks of 20 trials. The word lists were counterbalanced across blocks and conditions. Prior to starting each condition participants completed a practice version involving five trials.

Prospective Metacognitive Task

Participants had 20 seconds to study a list of 20 words presented on a computer screen. They were then asked to complete a distractor task (six maths questions). Once participants had completed the distractor task, they received instructions informing them that they would be presented with one word at a time and then to indicate if the word had been in the list that they had previously studied (a yes-no recognition test). This was the object-level component. Participants were also told that they would be asked to rate how confident they are about their answer on each trial. Following these instructions, a fixation cross appeared on the screen for 500ms followed by a word from the study list (n = 10) or a new previously-unpresented lure word (n = 10), these words appeared in a random order, and each was presented on the screen for 500ms. For each word, participants were then asked to rate how confident they were that they would make the correct decision about whether the word had been presented before. They made their judgements using a four-point scale, where one = not at all confident and four = extremely confident. Once they made their confidence judgement, they were then asked to indicate if the word had been previously presented in the study list.

Metacognitive accuracy for each condition was calculated in each participant by calculating a gamma correlation (Kruskal & Goodman, 1954) between word identification accuracy and confidence rating. Overall metacognitive accuracy was determined by calculating the average gamma between the two conditions as follows: (prospective gamma + retrospective gamma)/2. Gamma has been recommended by Nelson (1984), and Nelson, Narens, and Dunlosky (2004) and has been extensively used in research on metacognitive monitoring processes (e.g., Grainger et al., 2016; Sawyer et al., 2014; Williams et al., 2018). Use of gamma in the current study also serves to facilitate comparisons with other studies of metacognition in autism, which have almost exclusively employed gamma as the main dependent variable. Metacognitive accuracy ranges from − 1 to + 1 with scores of 0 indicating chance level accuracy, and large positive scores indicating good metacognitive accuracy.

Retrospective Metacognitive Task

This task followed the same procedure as the *prospective* condition described above, except participants were asked to indicate if the word had been previously presented in the study list *prior* to making their confidence judgment.

Autism-spectrum Quotient

The Autism-Spectrum Quotient (AQ) is a widely-used and well-validated self-report measure of autism traits, and it is considered to be a reliable measure of autism traits in both clinical and subclinical populations (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). The AQ presents participants with 50 individual statements (e.g., "I find social situations easy") and participants are asked to decide the extent to which they agree with each statement by responding on a 4-point Likert scale, ranging from "definitely agree" to "definitely disagree". Higher scores indicate more autism traits, with a maximum possible score of 50.

Reading the Mind in the Eyes Task

The Reading the Mind in the Eyes (RMIE) task is a widely used measure of mindreading ability among intellectually able adults, including those with autism (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). The adult version includes 36 photographs of the eye-region of the face. On each trial, participants are asked to pick one word from a selection of four to indicate what the person in the picture was thinking or feeling. Scores range from 0–36 with higher scores indicating better mindreading abilities. The task has been employed in over 250 studies, and shows good test-retest reliability (e.g., Fernández-Abascal, Cabello, Fernández-Berrocal, & Baron-Cohen, 2013), clearly distinguishes groups of participants with and without autism (e.g., Wilson et al., 2014), is associated with the number of autism traits shown by individuals in large population studies (e.g., Baron-Cohen et al., 2001), and is correlated with other measures of mindreading even after the influence of IQ is statistically controlled (e.g., Jones et al., 2018).

Statistical Analysis

Where results are predicted a priori on the basis of a solid theoretical foundation and/or previous empirical findings, reported significance values are for one-tailed tests. All other significance values are two-tailed. All correlational analysis used Pearson's correlation except where the data was skewed in which case Spearman's rho was used. Where t-tests were used, we report Cohen's d values as measures of effect size (≥ .0.20 = small effect, ≥ 0.50 = moderate effect; ≥ 0.80 = large effect; Cohen, 1969). Where ANOVAs were used, we report partial eta squared (η_p^2) values as measures of effect size (≥ .01 = small effect, ≥ .06 = moderate effect, ≥ .14 = large effect; Cohen 1969).

Mediation analysis for both prospective and retrospective metacognition was conducted using model 4 of the PROCESS macro for SPSS (Hayes, 2021). PROCESS utilises bootstrapping (10000 resamples) and is thus an appropriate technique to handle skewed data, which was the case for RMIE.

Experiment 1: Results

The mean overall gamma was .37, which was significantly different from zero, $t(81) = 13.40$, $p < .001$, $d = 1.48$. This indicates that participants were significantly above chance in their overall judgement-of-confidence accuracy, rating higher confidence for correct answers than for incorrect answers. The gamma scores were also significantly above chance in both the prospective condition, $t(81) = 12.01$, $p < .001$, $d = 1.33$, and the retrospective condition, $t(81) = 10.62$, $p < .001$, $d = 1.17$. Descriptive statistics for both metacognitive tasks are presented in Table 12. There was no significant difference between the two conditions for object-level or meta-level performance, but a borderline significant difference in confidence rating. Bivariate correlations among key variables are reported in Table 13.

Table 12

Experiment 1 Key Variables: Means, Standard Deviations (in brackets), and Inferential Statistics

Variable	Condition			
	Prospective	Retrospective	*T*	*p*
Object-level proportion correct	.74 (.12)	.75 (.13)	.82	.42
Average Confidence Rating	3.00 (.45)	2.92 (.49)	1.89	.06
Gamma	.35 (.26)	.40 (.34)	1.47	.15

Table 13

Experiment 1 Correlations between Key Variables

Variables	2	3	4	5
1. Prospective Gamma	.39***	.78***	-.33***	.31**
2. Retrospective Gamma		.88***	-.16^{m}	.26**
3. Overall Gamma			-.28**	.33***
4. Autism Quotient				-.26**
5. Reading the Mind in the Eyes				

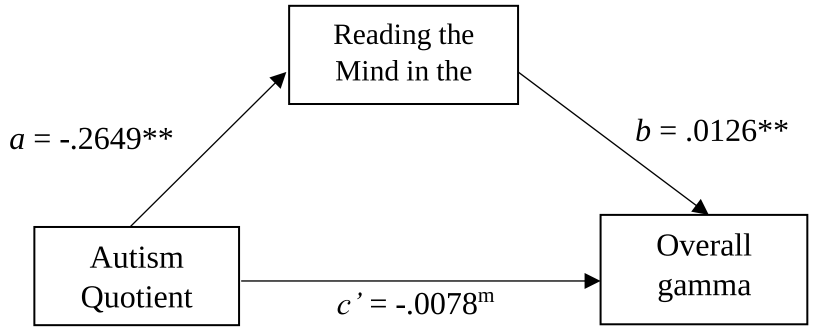
Note. * *p*<.05, ***p*<.01, ****p*<.001, ^{m} = Marginal (*p* <.08, two-tailed)

Mediation analysis

A series of mediation analyses were conducted to examine if the relationship between autism traits and metacognition was mediated by mindreading. Initial analysis showed that the effect of AQ on *overall* gamma was significant (B = -.011, SE = .004, t = -2.58, p <.01). The direct model was marginally significant (B = -.007, SE = .004, t = -1.79, p =.08) and the indirect effect (B = -.003, SE =.002) was significant (bootstrapped 95% confidence interval did not include zero, -.0088 to -.0001), indicating that the relationship between AQ score and overall gamma was significantly mediated by RMIE. The direct path was marginally significant and, therefore, RMIE accounts for some, but not all, of the relation between AQ and overall gamma.

Figure 8

Path Coefficients for the Mediating Effect of RMIE between AQ and Overall Gamma

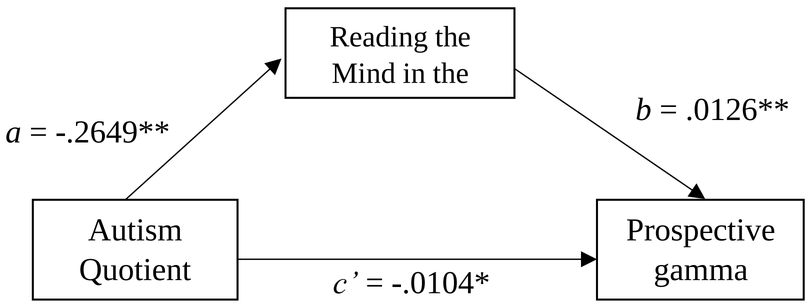


*p <.05, **p<.01, ***p<.001, m = ≤.08

Mediation analysis for *prospective* gamma showed that the total effect of AQ on prospective gamma was significant (*B* = -.014, *SE* = .004, *t* = -3.17, *p*<.01). The direct model remained significant (*B* = -.010, *SE* = .004, *t* = -2.37, *p* <.05) and the indirect effect (*B* = -.003, *SE* =.002) was significant (bootstrapped 95% confidence interval did not include zero, -.0082 to -.0004), indicating that the relationship between AQ total score and prospective gamma was significantly mediated by RMIE. The direct path was also significant and, therefore, mindreading accounts for some, but not all, of the relation between autism traits and gamma.

Figure 9

Path Coefficients for the Mediating Effect of RMIE between AQ and Prospective Gamma

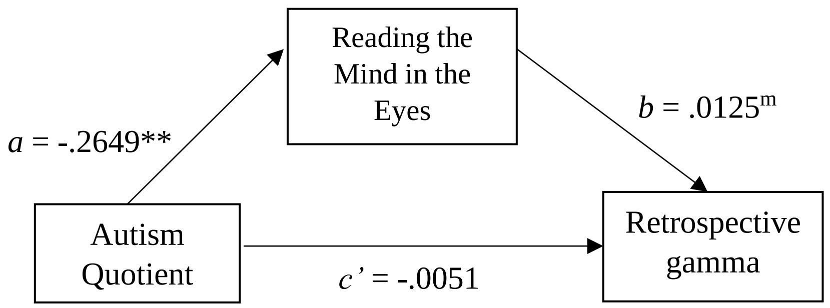


**p* <.05, ***p*<.01, ****p*<.001

Mediation analysis for *retrospective* gamma showed that the total effect of AQ on retrospective gamma was not significant (*B* = -.008, *SE* = .006, *t*=-2.76, *p* =.16). The direct model remained non-significant (*B* = -.005, *SE* = .006, *t*=-0.82, *p* =.41) and the indirect effect (*B* = -.003, *SE* =.003) was not significant (bootstrapped 95% confidence interval included zero, -.0099 to .0006), indicating that any relationship between AQ total score and retrospective gamma was not significantly mediated by RMIE.

Figure 10

Path Coefficients for the Mediating Effect of RMIE between AQ and Retrospective Gamma



*p <.05, **p<.01, ***p<.001, ^{m} = ≤.07

Controlling for object-level performance

The results from the correlation and mediation analyses suggest that mindreading is positively related to metacognition and mediates the link between metacognition and autism traits. However, to ensure that this apparent link between mindreading and metacognition isn't merely an artefact of cognitive ability underpinning performance on each type of task, it is important to control for cognitive level ability.

Subsequent partial correlation analyses showed that both AQ and RMIE scores were associated significantly with overall gamma score, even after controlling for object-level performance [AQ × overall gamma, $r(79) = -.21$, p <.05; RMIE × overall gamma, $r(79) = .23$, p <.05].

When each aspect of metacognition was examined separately, AQ and RMIE scores were associated significantly with gamma scores in the *prospective* condition, even after controlling for object-level performance [AQ × prospective gamma, $r(79) = -.30$, $p <.01$; RMIE × prospective gamma, $r(79) = .24$, $p <.05$]. However, after controlling for object-level performance, the RMIE × retrospective gamma was only borderline significant, $r(79) = .15$, $p =.09$. The AQ × gamma correlation was non-significant when object-level was controlled for, $r(79) = -.09$, $p =.23$.

Experiment 1: Discussion

As predicted our results showed that autism traits and mindreading were significantly related to metacognitive accuracy, with better mindreading performance and fewer autism traits being linked to better metacognitive accuracy. When we examined prospective and retrospective metacognition independently, both aspects were associated with mindreading ability and autism traits (though retrospective metacognition was only marginally significantly related to autism traits). Crucially, mindreading ability mediated the relation between metacognition and autism traits. However, when metacognition was broken down into *prospective* and *retrospective* aspects, mindreading ability only mediated the relation between autism traits and prospective metacognition. In keeping with these findings, the relations among overall, as well as prospective metacognition, on the one hand, and mindreading and autism traits, on the other hand, remained significant when controlling for object-level performance. However, the relationship between mindreading and retrospective metacognition reduced in strength and became only borderline significant when controlling for object-level performance. The relationship between autism traits and retrospective metacognition was no longer significant after controlling for object-level performance.

These results suggest that mindreading is more strongly and reliably related to prospective aspects of metacognition than to retrospective aspects of metacognition.

However, this difference *may* be one of degree, rather than absolute. That is, a relation between mindreading and retrospective aspects of metacognition may exist, but less strongly than the relation between mindreading and prospective aspects of metacognition.

Based on these results, it was predicted that the autism group would show significantly lower metacognitive accuracy than the typically developing group for overall metacognitive accuracy, as well as prospective metacognitive accuracy. It was also predicted that there would be a significant positive relationship between metacognitive accuracy (overall and prospective, at least) and mindreading. Finally, it was expected that there would be a significant negative association between metacognitive accuracy (overall and prospective, at least) and autism traits.

Experiment 2: Method

Participants

Twenty-nine adults with a diagnosis of autism (16 males) and 29 typically developing (15 males) adults took part in the current study. Four participants (2 ASD, 2 TD) were excluded from all analysis due to a lack of variation in confidence ratings in one or both metacognitive tasks, meaning it was not possible to calculate their gamma scores. This left twenty-seven autistic participants (14 males) and 27 typically developing participants (13 males). All of the participants in the autism group had received a formal diagnosis in accordance with established criteria (American Psychiatric Association, 2013; World Health Organization, 1993). Out of these 27 autistic participants, we were able to obtain Autism Diagnostic Observation Schedule (ADOS) scores for 16 of them, eight of whom scored seven or above. The ADOS is a semi-structured observational assessment of social, imagination and communication skills and a score of ≥ 7 is consistent with a diagnosis of autism (Lord et al., 2000). The mean ADOS total score among participants in the ASD group was 6.50 (SD = 3.85), which is below the clinical cut-off on this measure.

Details of participant characteristics can be seen in Table 14. Full Scale (FSIQ), Verbal (VIQ) and Performance (PIQ) IQ were assessed using the Wechsler Abbreviated Scale for Intelligence-II (Wechsler, 1999)[6]. All participants also completed the AQ as a measure of autism traits and RMIE as a measure of mindreading ability. Some participants also completed an animation task (Abell et al. 2000) as a measure of mindreading ability during one of our previous studies (24 in the autism group and 15 in the TD group). There were no significant differences between the autism and TD group in terms of age, FSIQ, VIQ, or PIQ, or mindreading ability. There were, however, expected between-group differences in number of autism traits (in line with their diagnostic status).

The study was ethically approved by the University's Research Ethics Committee (202016019942726630) and informed consent was obtained prior to commencing the tasks. All participants received payment of £10 per hour for their time, and all participants were provided with a written debrief following completion of all tasks, with the opportunity to contact the researchers should they have any further questions.

[6] Due to government restrictions imposed during the COVID-19 pandemic, 18 participants (7 ASD, 11 TD) completed three of the WASI subtests (vocabulary, matrix reasoning, and similarities) online. To obtain a full-scale IQ the block design subtest was given the same t-score as the matrix reasoning score.

Table 14

Experiment 2 Participant Characteristics: Means, Standard Deviations (in brackets), and Inferential Statistics for Autism and Typically Developing groups.

	Group				
	ASD (n = 27)	TD (n = 27)	*t*	*p*	*Cohen's d*
Age	33.47 (11.10)	31.34 (12.70)	0.66	.51	0.18
Full-scale IQ	104.00 (12.94)	101.30 (12.11)	0.79	.43	0.22
Performance IQ	102.63 (17.43)	100.30 (13.52)	0.55	.58	0.15
Verbal IQ	104.52 (9.63)	102.67 (10.89)	0.66	.51	0.18
Autism Quotient	32.96 (8.72)	16.41 (6.25)	8.02	<.001	2.18
Reading the Mind in the Eyes	26.37 (4.00)	27.00 (4.55)	0.54	.59	0.15
Animations[a]	5.25 (2.05)	5.80 (1.47)	0.97	.34	0.30

[a]scores for 24 autistic and 15 typically developing individuals

Materials, Procedure and Scoring

All participants completed the AQ, RMIE, and the *prospective* and *retrospective* judgements-of-confidence tasks all of which are described in Experiment 1. In addition to these tasks, some participants had completed an animations task (Abell et al. 2000) during one of our previous studies. This task has been widely used to assess mindreading abilities in both the general population and those diagnosed with autism (Wilson, 2021). During this task, participants are required to watch four short video clips of two triangles moving around. The clips are presented on a computer screen and, after watching each clip, participants are asked to describe what they think was happening in the video. Participants were allowed to watch each clip twice and responses are recorded using an audio recorder and later transcribed. Accurate responses required participants to attribute mental states, such as desire and intention, to the two triangles. Scores ranged from 0 to 2 for each clip, with higher scores indicating better mindreading abilities. Participants completed one practice trial prior to commencing the test trials.

Experiment 2: Results

Diagnostic Group Comparisons

Table 15 shows the means and standard deviations for each key variable. With regard to object-level performance, a 2 (Group: ASD/TD) × 2 (Condition: prospective/retrospective) mixed ANOVA was conducted. There was no significant main effect of condition, $F(1, 52) = 1.56$, $p = .22$, $\eta_p^2 = .03$ or group, $F(1, 52) = 0.14$, $p = .71$, $\eta_p^2 = .003$ and no significant interaction between group and condition $F(1, 52) = 0.07$, $p = .79$, $\eta_p^2 = .001$. This suggests that both the autism group and typically developing group performed to an equal level across both *prospective* and *retrospective* conditions in terms of number of words correctly identified. Thus, the two groups were very similar with respect to cognitive-/object-level ability.

In terms of confidence rating, a 2 (Group, ASD/TD) x 2 (Condition, prospective/retrospective) mixed ANOVA showed a significant main effect of condition, $F(1, 52) = 5.34$, $p = .03$, $\eta_p^2 = .09$, where participants placed higher ratings in the prospective condition than the retrospective condition (.12 (CI 95% .02 to .23), $p < .05$, $d = .30$). There was no significant main effect of group, $F(1, 52) = 0.54$, $p = .47$, $\eta_p^2 = .01$, and no significant interaction between group and condition $F(1, 52) = 0.18$, $p = .67$, $\eta_p^2 = .004$. Thus, confidence levels were equivalent across groups and groups showed similar patterns of confidence across conditions.

A 2 (Group, ASD/TD) × 2 (Condition, prospective/retrospective) mixed ANOVA showed for gamma that there was no significant main effect of condition, F(1, 52) = 2.34, $p = .13$, $\eta_p^2 = .04$, or group, $F(1, 52) = 0.00$, $p = .99$, $\eta_p^2 < .001$, and no significant group × condition interaction, $F(1, 52) = 0.06$, $p = .80$, $\eta_p^2 = .001$. This suggest that both groups performed to an equal level of metacognitive accuracy across both conditions.

Table 15

Experiment 2 Key Variables: Means and Standard Deviations (in brackets)

		ASD (n = 27)	TD (n = 27)
Overall	Object-level proportion correct	.77 (.12)	.76 (.12)
	Average Confidence Rating	2.92 (.49)	2.30 (.31)
	Gamma	.44 (.28)	.44 (.33)
Prospective	Object-level proportion correct	.79 (.12)	.76 (.12)
	Average Confidence Rating	2.96 (.56)	3.07 (.40)
	Gamma	.40 (.35)	.41 (.41)
Retrospective	Object-level proportion correct	.76 (.12)	.75 (.15)
	Average Confidence Rating	2.87 (.51)	2.93 (.30)
	Gamma	.48 (.23)	.47 (.32)

Mindreading Unmatched Comparison

Considering that there was no significant difference between the two groups in mindreading as measured by either the RMIE or the Animations tasks, and that mindreading ability is key to the theoretical debate, a group difference in mindreading ability was created and the metacognitive accuracy analysis was re-run. Initially, all participants who did not have an Animations task score were removed, leaving 24 autistic participants and 15 typically developing participants (9 males). We then removed the highest scoring autistic participants on the RMIE and Animations until there was a significant between-group difference on both mindreading measures [RMIE, $t(30)=-2.46$, $p = <.05$, $d = 0.87$; Animations, $t(30)=2.21$, $p = <.05$, $d = 0.78$. This left 17 (13 male) in the autism group. Groups remained matched for baseline in age, FSIQ, VIQ, PIQ (all *p*s $\geq .26$, all $d \leq 0.41$) but still differed on AQ, t(30) = 5.15, $p < .001$, $d = 1.82$.

Table 16 shows the means and standard deviations for each key variable. A 2 (Group: ASD/TD) × 2 (Condition: prospective/retrospective) mixed ANOVA showed that there was a moderate and borderline significant effect of group, $F(1, 30) = 3.07$, $p = .09$, $\eta_p^2 = .09$, but no main effect of condition, $F(1, 30) = 2.59$, $p = .12$, $\eta_p^2 = .08$. There was no significant group × condition interaction, $F(1, 30) = 0.59$, $p = .45$, $\eta_p^2 = .02$.

Table 16

Experiment 2 Key Variables: Means and Standard Deviations (in brackets) for groups unmatched on mindreading.

		ASD (n = 17)	TD (n = 15)
Overall	Object-level proportion correct	.73 (.11)	.79 (.10)
	Average Confidence Rating	2.80 (.52)	3.02 (.33)
	Gamma	.37 (.27)	.53 (.23)
Prospective	Object-level proportion correct	.75 (.11)	.79 (.11)
	Average Confidence Rating	2.81 (.58)	3.12 (.43)
	Gamma	.30 (.36)	.50 (.35)
Retrospective	Object-level proportion correct	.71 (.11)	.78 (.10)
	Average Confidence Rating	2.79 (.55)	2.93 (.31)
	Gamma	.45 (.27)	.56 (.24)

Correlations

To examine the relationship between autism traits (AQ), metacognition (gamma) and mindreading (RMIE) a series of correlational analyses was conducted. Given that there was no group difference in mindreading or metacognition, the analysis was conducted among both groups combined. All 27 ASD and 27 TD participants were included, given that the correlation analyses are essentially an individual differences approach. All results from the correlational analysis are presented in Table 17. In line with what was predicted, there was a significant association between Animations and overall gamma. However, this ceased to be significant when controlling for object-level performance $r(36) = -.01$, $p = .47$. There was also a marginally significant association between overall gamma and RMIE, but this became non-significant when controlling for object-level performance $r(51) = .11$, $p = .23$.

When broken down there was a significant association between prospective gamma and both RMIE and Animations. The RMIE × prospective gamma was no longer significant when controlling for object-level performance $r(51) = .07$, $p = .31$. The Animations × prospective gamma was marginally significant, $r(36) = .25$, $p = 06$ after controlling for object-level performance. There was also a significant association between Animations and gamma in the retrospective condition, this was no longer significant when controlling for object-level performance $r(36) = -.20$, $p = .11$.

Analysis also confirmed that there were no significant differences between the autism group and the typically developing group for any of the correlations according to Fisher's Z test, all zs < .94, all *p*s > .35, except for the retrospective gamma × AQ, $z = 1.74$, $p = 04$. Therefore, this association was analysed in each group separately. Among ASD participants, the retrospective gamma × AQ was not significant, $r(27) = .20$, $p = .15$. Among the TD participants the retrospective gamma x AQ was marginally significant,

$r(27) = -.29$, $p = .07$, which was no longer significant when object-level performance was controlled for, $r(24) = -.24$, $p = .12$

Table 17

Experiment 2: Correlations between Key Variables

	Variables	2	3	4	5	6
1.	Prospective Gamma	.45***	.87***	.12	.23*	.46**
2.	Retrospective Gamma		.82***	-.02	.17	.31*
3.	Overall Gamma			.07	$.20^{m}$	.45**
4.	Autism-Spectrum Quotient				$-.20^{m}$	$.23^{m}$
5.	Reading the Mind in the Eyes					.31*
6.	Animations					

Note. * $p<.05$, ** $p<.01$, *** $p<.001$, ^{m} = Marginal ($p<.08$)

Experiment 2: Discussion

The results from Experiment 2 revealed a significant association between metacognitive accuracy and mindreading ability. However, after controlling for object-level performance, only the association between prospective metacognition and mindreading performance on the animations remained (borderline) significant. Interestingly, there was no significant between-group difference in overall, prospective, or retrospective metacognitive accuracy. This suggest that the autism group were as accurate in their metacognitive judgements as the typically developing group. While this was contrary to predictions, it is also unsurprising considering that there was no difference in mindreading ability between the two groups. The one system account only predicts diminished metacognitive accuracy among groups with diminished mindreading ability. Somewhat (but not entirely) in line with the one system account, there was a borderline significant between-group difference in metacognitive accuracy when a between-group difference in mindreading ability was created. Although this difference was only borderline significant in this reduced sample (p = .09), it was associated with a moderate effect size (d = 0.64) and suggests that when mindreading is impaired then a metacognitive difficulty of similar magnitude is apparent.

General Discussion

To our knowledge, this is the first study to examine both *prospective* and *retrospective* metacognition in relation to autism and mindreading among adults within a common sample using the same paradigm. Furthermore, whilst prospective metacognition has been examined in relation to autism and mindreading using feeling-of-knowing and judgement-of-learning tasks, this is the first study to examine accuracy of *prospective* confidence judgements in relation to autism and mindreading.

As expected, the results from Experiment 1 showed that metacognitive accuracy was significantly associated with mindreading in a student sample, indicating that the better an individual's mindreading ability the more accurate their metacognitive judgements. This is in line with the one-system account but contrasts with the findings from the studies presented in chapter three and four of this thesis. This link persisted regardless of the type of metacognitive judgement, *prospective* or *retrospective*, suggesting that accuracy of confidence judgements are linked to mindreading regardless of whether the judgement is about past or a future performance. Importantly, these associations remained significant (or borderline significant in the case of retrospective judgements) after object-level cognitive ability was controlled. Furthermore, mindreading ability was found to mediate the link between metacognition (overall and prospective metacognition, at least) and autism traits. Thus, these findings add further support to previous research into *retrospective* judgements-of-confidence accuracy (Grainger et al., 2016; McMahon et al., 2016; Nicholson et al., 2019; Nicholson et al., 2020; Wilkinson et al., 2010; Williams et al., 2018) and provide new evidence for the accuracy of confidence judgements about future performance.

These findings were only partially replicated among a sample of autistic and typically developing adults in Experiment 2. While bivariate correlation analysis suggested the same links among metacognition, mindreading, and autism traits seen in Experiment 1, results from partial correlation analyses in Experiment 2 did not replicate those results from Experiment 1. Specifically, all but one correlation (the prospective metacognition × Animations task correlation) ceased to be statistically significant after controlling for the effects of object-level performance. Of course, the sample size in Experiment was not as large as in Experiment 1 and so Experiment 2 had lower power to detect true relations among these measures. Moreover, controlling for object-level performance is a conservative/stringent approach, given that cognitive and metacognitive levels influence each other in a bidirectional manner (Nelson & Narens, 1990). Therefore, controlling for object-level performance will mask some of the independent relation between metacognitive accuracy and mindreading. Nonetheless, the results from Experiment 2 were clearly not straightforwardly reconcilable with the results from Experiment 1.

More puzzling than the results from the correlation analyses in Experiment, was the absence of a between-group difference in metacognitive accuracy. In Experiment 1, metacognitive accuracy was associated significantly with autism traits and this association was mediated by mindreading ability. Therefore, we expected autistic participants to manifest diminished metacognitive accuracy in Experiment 2, on the basis that they would show high autism traits (by definition), as well as characteristic difficulties with mindreading. In line with simulation theory and theory-theory the predicted difficulties with metacognition were not observed. This confirms the results found from the studies presented in Chapters three and four, however, note that the characteristic difficulties with mindreading were also not observed in this sample. When the participant groups were "unmatched" for mindreading ability, however, between-group differences in metacognitive accuracy were clearer and moderate in size (d = 0.64), although not quite significant according to one-tailed analysis (p = .09) in the reduced sample of 17 autistic and 15 typically developing participants.

The between-group differences in Experiment 2 are difficult to interpret and conclusions can only be made tentatively. This points to a wider issue with sampling that the field needs to address. Only if the groups in case-control studies are broadly representative of their respective populations can meaningful conclusions from experimental results be drawn. What counts as "representative" isn't straightforward to define and depends to some extent on the topic being studied. In the current study, it was important that the autism group did not show the difficulties with mindreading that are characteristic of autism, and which are observed in most studies. Given the theoretical link between mindreading and metacognition, we would not predict a difficulty with metacognition in a sample of autistic participants who did not manifest a difficulty with mindreading. Therefore, it was important to un-match the samples in the current study to artificially create a group difference in mindreading. The difficulty with this approach, of course, is that it reduces sample size and, thus, statistical power to detect group differences in the experimental variable of interest, namely metacognitive accuracy.

Moreover, even after this un-matching, there was some evidence that the autistic sample in the current study was somewhat unrepresentative/atypical in composition.

Among those participants who completed the ADOS, only half scored above the clinical cut-off score of 7, and the group average was below the clinical threshold on the measure. This does not necessarily mean that the autistic participants who scored below 7 were misdiagnosed with autism; sensitivity of the ADOS is far from perfect, meaning that a relatively substantial proportion of individuals who are judged by clinicians to warrant an autism diagnosis do not score above 7 on ADOS (Risi et al., 2006). However, it is equally true that specificity of ADOS is far from perfect, with approximately 29% of children judged as non-spectrum by expert clinicians scoring above threshold on the ADOS (Risi et al., 2006). Ultimately, performance on the ADOS or any other measure doesn't over-rule the diagnostic judgement of an expert clinician. However, it *would* be sensible for future studies of metacognition in autism to include a particular level of observable autism severity as an inclusion criterion.

Turning our attention to the analysis for autism traits, Experiment 1 revealed that there was a significant association between the number of autism traits and *overall* metacognitive accuracy, with more autism traits being linked to poorer accuracy (replicating Carpenter et al., 2019 – Chapter three but contrasting with Chapter four). Further examination of our results showed that autism traits are not only related to metacognition but that the relationship is mediated by mindreading**.** This mediation is exactly what would be expected from a one-system point of view, given that the one-system account argues that metacognition is diminished in autism due to difficulties in mindreading.

When the types of meta-level judgements were examined separately this mediation persisted in the *prospective* condition but not the *retrospective* condition. This suggests that different mechanisms may be involved when making prospective confidence judgements in comparison to retrospective confidence judgment. Interestingly, the relationship between autism traits and prospective metacognitive performance is only partially mediated by mindreading abilities, indicating that autism traits independent of mindreading ability have some role to play in prospective judgements-of-confidence accuracy. Research has shown that individuals diagnosed with autism have significant impairments in multiple areas of executive functioning (Corbett, Constantine, Hendren, Rocke & Ozonoff, 2009; Berenguer, Roselló, Colomer, Baixauli, & Miranda, 2018), and executive functioning has been shown to play some part in metacognitive decisions (Fernandez-Duque, Baird, and Posner, 2000). Therefore, it is possible that the difficulties in executive functioning linked to autism also plays a role in autistic individual's *prospective* metacognitive abilities. Despite the findings regarding autism traits in Experiment 1, we did not replicate the association between autism traits and metacognitive accuracy in Experiment 2.

Overall, this study shows a clear relationship between mindreading and metacognitive accuracy, regardless of whether the judgement is about a past or future event. This is in line with the 'metacognition-is-prior' and the 'one-system' account predictions and contrasts with the 'two-system account'. The study also showed that autistic adults are able to make metacognitive judgement to an equal level of accuracy as typically developing people. This supports the predictions made by simulation theory and theory-theory, although when there is a difference between groups in mindreading ability the difference in metacognitive accuracy is large and marginally significant. This suggests that only autistic people with significant mindreading difficulties will have difficulties with metacognition however further research is required to confirm this conclusion.

Beyond the theoretical debate, metacognition has been shown to relate to various aspects of everyday functioning including self-regulation (behavioural and cognitive; Nelson & Narens, 1990), motivation (Metcalfe & Finn, 2008), and decision making (Yeung & Summerfield, 2012) and can have an impact on numerous life situations including academic achievement, employability, and eyewitness testimony (Hartwig & Dunlowsky, 2012; Veenman & Spaans., 2005; Zaragoza & Mitchell 1996). Therefore, the results from this study have clear clinical and practical implications for individuals with diminished mindreading ability and may shape the support recommended/provided such individuals.

Chapter Six: Examining Metacognition for Perceptual and Semantic Tasks in Relation to Mindreading and Autism using a Post-Decision Wagering Paradigm

Abstract

Background: There is evidence to suggest that metacognitive accuracy may vary in accordance with the requirements of the cognitive-level task. Considering that the study presented in chapter three is the only study to have used post-decision wagering. The study presented in this chapter aims to expand these findings by employing the same post-decision wagering paradigm but with a semantic cognitive-level task, as well as the perceptual discrimination task used in chapters three and four. ***Method*****:** In Experiment 1, 129 students participated in both the perceptual and semantic wagering task. The perceptual task was the same as that used in Experiment 1 of chapter three. For the semantic task participants were presented with a word at the top of the screen and were required to select the synonym from two words presented below it. Following their decision participants made their meta-level judgement by placing a wager on their answer being correct. For experiment 2, 26 autistic and 25 typically developing adults completed the same task as that presented in Experiment 1. ***Results:*** There was a significant positive association between mindreading and metacognitive accuracy on both the perceptual and semantic task in Experiment 1. This association was replicated in Experiment 2 for metacognitive accuracy on the semantic task but not the perceptual task. There was also a significant negative association between autism traits and metacognitive accuracy on the perceptual task in Experiment 1, replicating the findings in chapter three. In line with the results from previous chapters there was no significant between group difference in meta-cognitive accuracy. There was also no significant between group difference in meta-level reaction time, this contrasts with the findings from chapters three and four.

The ability/abilities to attribute mental states to *oneself* (metacognition) and *others* (mindreading) is central to how we make sense of and predict our own and others behaviour (Flavell, 1979). Impairments in one or both of these skills can have significant implications for everyday functioning and life chances. On the one hand, mindreading ability has been linked to difficulties with social interaction (Frith et al., 1994; Tager-Flusberg, 1999). On the other hand, metacognition has been linked to difficulties with decision making, planning, and even academic success (Tobias & Everson, 2009). Given that both mindreading and metacognition require the ability to attribute metal states, it could be predicted that these two abilities are linked and rely on the same underlying processes (e.g., Carruthers, 2009). Indeed, the idea that the two abilities are linked was inherent in Premack and Woodruff's original definition of "theory of mind" as the ability to attribute mental states to oneself *and* others to explain and predict behaviour (Premack & Woodruff, 1978).

One way to examine this potential link is to examine metacognitive ability among a population that is known to have difficulties with mindreading. One such group of people are those diagnosed autism spectrum disorder (ASD, autism). Autism is a neurodevelopmental condition characterised by social-communication difficulties, and restricted and repetitive behaviours and interests (American Psychiatric Association, 2013). A key feature of autism is difficulties with mindreading, at least at a cognitive level (Brunsdon & Happé, 2014). Therefore, if mindreading and metacognition are related as some suggest, autistic people will also have diminished metacognition. To find deminished metacognition would potentially pose a challenge to those theorists who claim that metacognition and mindreading rely on different underlying cognitive mechanisms (e.g., Nichols & Stich, 2003), and those who claim that introspective access to one's own mental states is direct and non-inferential (e.g., Goldman, 2006).

In addition to the theoretical debate, examining metacognition among autistic people is key to understanding how autistic people process the world and make sense of their experiences. Moreover, if autistic people struggle to accurately monitor their own cognitive processes, then this may impact on their academic performance, ability to obtain and maintain employment, as well as make decisions in line with their true cognitive ability (Dunlosky & Metcalf, 2009; Hartwig & Dunlowsky, 2012; Hendricks, 2010; Shattuck et al., 2007; Veenman & Spaans, 2005).

Although there are many ways to examine metacognitive accuracy, the majority of studies into metacognition and autism have used the classic judgements-of-confidence task. Judgements-of-confidence tasks require participants to make a cognitive-level (object-level) decision, such as deciding which of two images has the most dots, and then to rate how confident they are that they made the correct decision. The closer the correspondence between cognitive performance and metacognitive judgements of cognitive performance, the better a person's metacognitive accuracy. Findings from these studies indicate that judgments of confidence among autistic children appears to be diminished across a variety of cognitive (object-level) domains, including general knowledge, emotion recognition, and perceptual decisions (Grainger et al., 2016; McMahon, 2016; Nicholson et al., 2020; Williams et al., 2018).

Findings among autistic adults, however, appear to be slightly more complex (See Chapter two). One possible explanation for this is the domain of cognition about which metacognitive judgements are being made. For, example, accuracy of confidence judgments about perceptual discrimination performance have been shown to be diminished among autistic adults (Nicholson et al 2019), but intact when such judgements are made about general knowledge, emotion recognition, and *some* aspects of memory (Cooper et al., 2016; Maras et al 2020; Sawyer et al 2014).

There is reason to suspect that the use of different object-level tasks may go some way to explaining the differences found in metacognitive ability among autistic adults. Firstly, research has shown that different brain mechanisms are involved in metacognitive judgements on different object-level tasks (Baird et al., 2013). Secondly, studies into judgement-of-confidence accuracy and autism have suggested that metamemory may be impaired among autistic adults only when certain encoding processes are involved (Cooper et al., 2016).

Up until now, only two studies have examined the accuracy of metacognitive judgements in response to perceptual decisions among autistic adults. Nicholson et al. (2019) found that judgement-of-confidence accuracy was diminished among autistic adults during a perceptual discrimination (object-level) task. Nicholson et al. also found that mindreading ability was associated significantly with judgement-of-confidence accuracy in both groups.

Carpenter et al. (2019) employed the same perceptual discrimination task as used by Nicholson et al. However, instead of using the classic judgements-of-confidence procedure to test metacognitive ability, Carpenter et al. employed an alternative (behavioural) means of measuring metacognition, namely post-decision wagering (PDW). PDW requires participants to place a wager on their object-level performance being correct rather than rating their confidence. If their object-level answer is correct, then the participant will win what they have wagered, but if they are incorrect, they will lose what has been wagered. So high wagers for correct answers and low wagers for incorrect answers would indicate good metacognitive accuracy. Evidence shows that PDW is good measure of metacognitive awareness among children and adults, with wagering reflecting participant's level of uncertainty as well as the number of correct and incorrect answers (Persaud et al., 2007; Ruffman, Garnham, Import, & Connolly, 2001).

Moreover, evidence indicates that PDW is as sensitive and as confidence ratings when it comes to measuring metacognitive accuracy, providing that risk aversion is accounted for (Dienes & Seth, 2010).

Unlike Nicholson et al., Carpenter et al. found metacognitive accuracy to be undiminished in a group of autistic participants compared to a group of typically developing adults (Exp. 2). They also found metacognitive accuracy to be unrelated to mindreading ability in a sample of students (Exp. 1) as well as in a group of autistic adults and their typically developing counterparts (Exp.2). However, Carpenter et al. also investigated reaction times as a means of assessing metacognitive ability and found a different pattern of results. Specifically, they found that autistic participants were slower at making metacognitive judgements in comparison to typically developing participants, even when controlling for the speed of their cognitive-level judgements. This implies that typically developing participants were able to access metacognitive processing and interpret it quicker, and thus arrive at a wagering decision with relative ease in comparison to the autistic participants. Crucially, this difference was not due to decision-making, motor co-ordination, or general speed of processing. These results were reinforced by the finding that the better the participants mindreading ability the quicker their meta-level judgements, even after controlling for cognitive-level speed.

Carpenter et al.'s (2019) results suggest that differences across studies in findings regarding metacognitive performance in autism might be explained not only by the kind of object-level measure used, but also the means of assessing metacognition. Therefore, it is important to extend research into metacognition in autism to measures that probe metacognitive awareness in alternative ways and to avoid the over-reliance on one measure (Cichoń et al., 2018). Thus, the current study employed a post-decision wagering paradigm to examine metacognitive accuracy in relation to mindreading and autism. It also extends previous research by examining the accuracy of metacognitive awareness for perceptual and sematic decisions among a common group of participants.

In line with Carpenter et al.'s (2019) study, an individual differences approach (Exp. 1) and a case control approach (Exp. 2). Using these two approaches is not only legitimate but is also recommended when it comes conducting autism research (Landry & Chouinard, 2016). Firstly, autism traits have been found to span both clinical and subclinical population making it reasonable to explore such traits within a typically developing sample. These milder or subclinical traits are often referred to as the broad autism phenotype (BAP). Secondly, it allows for a larger sample sizes and hence greater power to detect true effect, something that case control studies are often accused of lacking. It also allows for key variables to be examined independently of any comorbid conditions, giving a clearer picture of the relationships between the variables (Chouinard et al., 2016). That is not to say that case control studies are irrelevant, indeed such studies are required considering that there are characteristics unique to autism that are not found within the BAP (Landry & Chouinard, 2016). Moreover, whilst the BAP resembles autism it does not necessarily have the same underlying causes as those traits found among autistic individuals (Constantino & Todd, 2003; Landry & Chouinard, 2016; Williams et al., 2017). Therefore, the individual differences approach will be used to complement the case control approach within this study.

Experiment 1: Method

Participants

One hundred and thirty-five students from a university in the southeast of England took part in the experiment. Six participants were excluded from all analysis due to a lack of variation in wagering responses in one or both metacognitive tasks, meaning it was not possible to calculate their gamma scores. This left one hundred and twenty-nine participants in total (11 males, and one participant did not disclose their gender), with a mean age of 19.76 years (*SD* = 3.18; range = 18–44). Participants received course credits in partial fulfilment of their degree. The study was ethically approved by the University's Research Ethics Committee (202015984328836586) and informed consent was obtained prior to commencing the tasks. All participants were given a written debrief following completion of the tasks and could contact the researchers for further debrief if required.

Materials, procedure and scoring

This was an online study implemented using Psychopy 3 (Peirce et al., 2019) and Qualtrics software (2020), for the metacognitive tasks and the questionnaires, respectively. Participants were instructed to use their own laptop or desktop in a quiet space without any distractions. Participants completed two metacognitive tasks, a perceptual wagering task and a semantic wagering task. The order in which these tasks were completed were counter-balanced across participants. Participants also completed the Autism Quotient (*M* = 19.52, *SD* = 7.37; Baron-Cohen et al. 2001a) and Reading the Mind in the Eyes (*M* = 24.44, *SD* = 4.91; Baron-Cohen et al. 2001b), as a measure of autism traits and mindreading ability, respectively. They also completed the Balloon Analogue Risk Task (*M* = 24.13, *SD* = 11.03; Lejuez et al., 2002) as a measure of risk propensity.

Perceptual Wagering Task

This task was similar to the dots wagering task used by Carpenter, Williams and Nicholson., (2019). There were two phases to the task, the *perceptual judgement phase* and the *wagering phase* (see Figure 11).

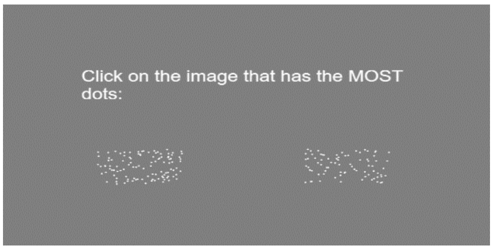
Perceptual judgement (object-level) phase. During this phase, participants were shown a series of images made up of dots (50 trials) on a computer screen. Participants were presented with two images on each trial and asked to identify which image had the most dots by clicking on the image using the mouse. The difficulty of the perceptual discrimination varied randomly across trials. Trial difficulty was operationalised in terms of the relative difference in the number of dots present in each of the two images. For example, a trial on which image A had 95 dots and image B had 125 dots (a proportional difference of .24) would be easier than a trial on which image A had 114 dots and image B had 120 dots (a proportional difference of .05). On each trial, participants had four seconds to make their judgement. If they had not made their judgement after five seconds, the programme moved on to the next trial and the trial was counted as a "miss". The proportion of trials on which a correct perceptual discrimination was made was used as one dependent variable. The average time it took participants to make their judgment was used as the second dependent variable. The quicker the discrimination response, the easier participants found it to make their judgement.

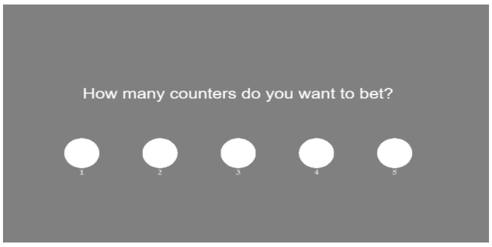
Wagering Phase. On each trial, after making their perceptual judgement, participants were asked to place a wager on their answer. Participants indicated how many counters they wished to bet by using a five-point scale. Participants were informed at the beginning of the task that if they correctly identified the image with the most dots then (a) they won back the counters they wagered plus one for every counter they wagered and (b) got to keep the counters that they did not bet. So, for example, if a participant bet three counters and their answer was correct they received the three counters they bet, plus three more and got to keep the two they had left over (thus, eight counters in total). If the participant bet three counters and their answer was incorrect, they lost the counters they wagered but got to keep the counters they had left over (i.e., if they bet three of the five counters they received two counters in total). Participants were not told how much they had won until all trials were complete. Participants could win up to 500 counters. Participants were informed that the top three people with the most counters at the end of the study would receive a prize (1st = prize worth £20, 2nd = prize worth £10 and 3rd = prize worth £5). Prior to commencing the trials participants completed five practice trials.

"Meta-level" performance was indexed in each participant by calculating a gamma correlation (Kruskal and Goodman 1954) between perceptual discrimination accuracy and number of counters wagered, providing a measure of metacognitive accuracy. This measure has been recommended by Nelson (1984), and Nelson et al. (2004) and has been extensively used in research on metacognitive monitoring processes (e.g. Grainger et al. 2016; Sawyer et al. 2014; Williams et al. 2018). Use of gamma in the current study also serves to facilitate comparisons with other studies of metacognition in ASD, which have almost exclusively employed gamma as the main dependent variable. Metacognitive accuracy ranges from − 1 to + 1 with scores of 0 indicating chance level accuracy, and large positive scores indicating good metacognitive accuracy. "Meta-level" performance was also indexed by the average time it took for participants to place their bet across trials. The quicker the wagering response, the easier participants found it to make their judgement. One participant was excluded from all analysis because there was no variation in their wagers across trials and so a gamma score could not be calculated (leaving n = 39).

Figure 11

Example Trial in the Perceptual Wagering Task





Perceptual Judgement Phase - Participants had four seconds to select the image with the most dots.

Wagering Phase - Participants took as long as they needed to place a wager on their perceptual judgement being correct.

Semantic Wagering Task

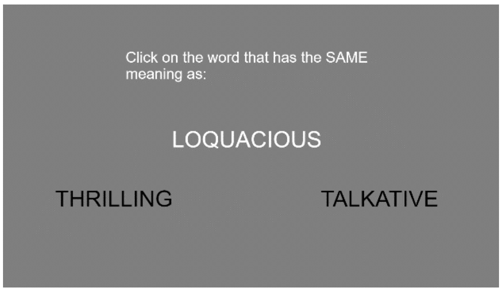
As with the perceptual wagering task there were two phases to the semantic wagering task, the *semantic judgement phase* and the *wagering phase* (see Figure 12).

Semantic judgement (object-level) phase. During this phase, participants were presented with series of words (50 trials) on a computer screen. Each trial consisted of one word displayed in the middle of the screen (main word), and two words below, one of which was a synonym of the top word and the other being a lure word. The words were chosen from the MRC Psycholinguistic Database (Coltheart, 1981). Participants were asked to use a mouse to identify which out of the two bottom words was a synonym of the main word. They had five seconds to make their judgement. The difficulty of the semantic discrimination varied randomly across trials. Trial difficulty was operationalised in terms of a) Kucera-Francis written frequency of the main word and b) the synonym/lure word. The lower the KF written frequency of the words, the more difficult the trial. Synonym and lure words were matched closely for Kucera-Francis written frequency, word length, and starting letter. The higher the Kucera-Francis written frequency for the main word, synonym word and lure word combined on each trial, the easier the trial was considered to be. For example, a trial with a total KF of 643 was considered easier than a trial with a total KF of 6, see Appendix 2 for a list of the words included in each trial and their KF. On each trial, participants had four seconds to make their judgement. If they had not made their judgement after four seconds, the programme moved on to the next trial and the trial was counted as a "miss". The proportion of trials on which a correct semantic discrimination was made was used as one dependent variable. The average time it took participants to make their judgment was used as the second dependent variable. As with the perceptual wagering task, the quicker the discrimination response, the easier participants found it to make their judgement.

Wagering Phase. The wagering phase of the sematic wagering task was the same as the wagering phase in the perceptual wagering task. Participants also completed five practice trials before commencing the main trials.

Figure 12

Example Trial in the Semantic Wagering Task



Semantic Judgement Phase - Participants had four seconds to select the correct synonym.

Wagering Phase - Participants took as long as they needed to place a wager on their semantic judgement being correct.

Autism-Spectrum Quotient

The Autism-spectrum Quotient (AQ) is a widely-used and well-validated self-report measure of ASD traits. It is considered to be a reliable measure of ASD traits in both clinical and subclinical populations (Baron-Cohen et al. 2001a). The AQ presents participants with 50 individual statements (e.g., "I find social situations easy") and participants were asked to decide the extent to which they agreed with each statement by responding on a 4-point Likert scale, ranging from "definitely agree" to "definitely disagree". Higher scores indicate more ASD traits, with a maximum possible score of 50.

Reading the Mind in the Eyes Task

The Reading the Mind in the Eyes Task (RMIE) task is a widely used measure of mindreading among intellectually able adults, including those with ASD (Baron-Cohen et al. 2001b). The task involves looking at photographs of eyes and deciding what the person in the picture is feeling. Participants were presented with 36 eye stimuli and were required to select an emotion that best described what the person in the picture may be feeling out of four possible emotions. Scores ranged from 0 to 36 with higher scores indicating better mindreading abilities.

Balloon Analogue Risk Task

The Balloon Analogue Risk Task (BART) is a computer-based task designed to measure risk propensity (Lejuez et al., 2002). In this task participants were required to inflate a computer-simulated balloon by pressing the space bar. In the current study, participants earned virtual money with each pump, which was later converted into points (£1 = 1 point) and added on to their score on the wagering tasks. The amount earned in each trial was displayed on the screen with the total amount earned being presented throughout.

When the balloon was pumped up too much, resulting in it exploding, participants did not gain anything for that trial. Participants were able to cease pumping the balloon at any point and bank the gains earned for that trial adding it to the total earnings. There were 20 trials in total.

Statistical Analysis

Results that were predicted a priori on the basis of a solid theoretical foundation and/or previous empirical findings, reported significance values are for one-tailed tests. All other significance values are two-tailed. Where t-tests were used, we report Cohen's *d* values as measures of effect size (≥ .0.20 = small effect, ≥ 0.50 = moderate effect; ≥ 0.80 = large effect; Cohen 1969). All correlational analysis used Pearson's correlation except where the data was skewed in which case Spearman's rho was used. Where t-tests were used, we report Cohen's *d* values as measures of effect size (≥ .0.20 = small effect, ≥ 0.50 = moderate effect; ≥ 0.80 = large effect; Cohen, 1969). Where ANOVAs were used, we report partial eta squared (η_p^2) values as measures of effect size (≥ .01 = small effect, ≥ .06 = moderate effect, ≥ .14 = large effect; Cohen 1969).

Experiment 1: Results

The overall gamma was .38, which was significantly different from zero, $t(128) = 19.68$, $p < .001$, $d = 1.65$. This indicates that participants were significantly above chance in their overall judgement-of-confidence accuracy, rating higher confidence for correct answers than for incorrect answers. Gamma scores were significantly above chance in both the semantic condition, $t(128) = 9.95$, $p < .001$, $d = 2.18$, and the perceptual condition, $t(128) = 24.34$, $p < .001$, $d = 0.88$. Descriptive statistics for both metacognitive tasks are presented in Table 18. There was no significant difference between the two conditions for object-level, $t(128) = 1.59$, $p = .11$, $d = 0.11$. There was, however a significant difference in meta-level performance, $t(128) = 7.25$, $p < .001$, $d = 0.67$, with higher gamma in the semantic condition. There was also a significant difference in wager, $t(128) = 5.76$, $p < .001$, $d = 0.59$, with participants placing higher ratings for the semantic condition compared to the perceptual condition. Finally, participants were quicker at making meta-level judgements in the perceptual condition compared to the semantic condition, $t(128) = -5.58$, $p < .001$, $d = 0.53$.

Table 18

Experiment 1 Key Variables: Means, Standard Deviations (in brackets)

Variable	Condition	
	Perceptual	Semantic
Object-level proportion correct	.66 (.09)	.67 (.09)
Missed trials	1.84 (2.88	2.92 (4.24)
Object-level reaction time (s)	2.20 (0.42)	2.71 (0.37)
Counters wagered	2.76 (0.80)	3.19 (0.72)
Wagering reaction time (s)	1.23 (0.59)	1.60 (0.77)
‘Meta-level’ Gamma	.28 (.32)	.48 (.22)

Correlations

The results from the correlational analysis are presented in Table 19. In line with what was predicted, RMIE and AQ scores were significantly associated with overall gamma, with higher RMIE score and lower AQ scores being associated with higher gamma score. The RMIE × overall gamma association remained significant when controlling for object-level performance, $r_s(126) = .29$, $p <.001$. AQ × overall gamma remained borderline significant when object-level performance was controlled for, $r_s(126) = -.13$, $p = .08$.

When each condition was examined separately, RMIE scores were associated significantly with gamma scores in both the *perceptual* and *semantic* condition. These associations each remained significant when controlling for object-level performance [RMIE x perceptual gamma, $r_s(126) = .18$, $p <.05$; RMIE × semantic gamma $r_s(126) = .30$, $p <.001$]. Fisher's Z revealed that there was no significant between-condition (semantic/perceptual) difference in the strength of correlations between RMIE and gamma, Z = -.05, $p = .32$.

AQ was also associated significantly with gamma scores in the *perceptual* condition, but not the *semantic* condition. However, Fisher's Z showed that magnitude of the AQ × semantic gamma correlation was non-significantly smaller than the AQ × perceptual gamma, Z = -1.29, $p = .10$. Importantly, the AQ x *perceptual* gamma remained significant when object-level performance was controlled for $r_s(126) = -.15$, $p <.05$.

In terms of wagering reaction time there was a significant correlation between overall meta-level wagering reaction time and AQ, this was borderline significant when controlling for object-level reaction time, $r_s(126) = -.14$, $p =.06$. There was also a significant correlation between AQ and wagering reaction time for the *perceptual* task but not the *semantic* task. The AQ × wagering reaction time for the *perceptual* task became marginal when object-level reaction time was controlled for, $r_s(126) = -.14$, $p =.06$.

Table 19

Experiment 1: Correlations between Key Variables

Variables	2	3	4	5	6	7	8	9
1. Autism Quotient	-.16*	-.15*	-.18*	-.01	-.11	-.11[m]	-.16*	-.07
2. Reading the Mind in the Eyes		.20*	.07	.25**	.00	.30***	.05	.21*
3. Perceptual Gamma			.04	.25**	.01	.88***	.01	.15[m]
4. Perceptual wagering reaction time				-.01	.57***	.03	.83***	.00
5. Semantic Gamma					-.01	.65***	-.04	.20*
6. Semantic wagering reaction time						-.02	.91***	-.02
7. Overall Gamma							.00	.22*
8. Overall wagering reaction time								.01
9. Balloon Analogue Risk Task								

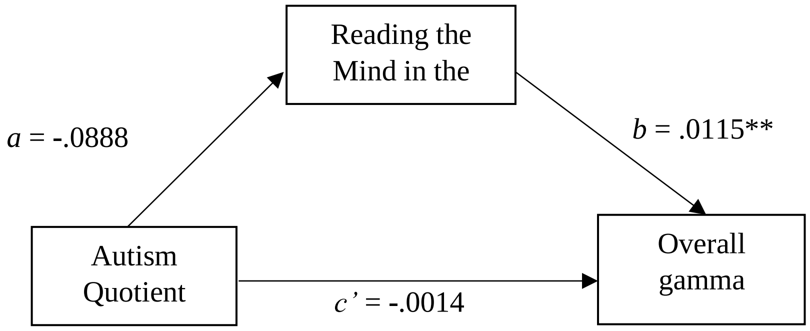
Note. * p<.05, **p<.01, *** p<.001, m = ≤.09

Mediation analysis

Given that there was significant/marginally significant association between mindreading, autism traits and overall metacognition, a mediation analysis was conducted to examine if the relationship between autism and metacognition was mediated by mindreading. The analysis showed that the effect of AQ on overall gamma was not significant (B = -.002, SE = .003, *t* = -.93, *p* = .35). The direct model was not significant (B = -.001, SE = .003, *t* = -.55, *p* =.58) and the indirect effect was not significant (B = -.001, SE <.001; (bootstrapped 95% confidence interval included zero, -.0029 to .0004), indicating that any relationship between AQ score and overall gamma was not mediated by RMIE.

Figure 13

Path Coefficients for the Mediating Effect of RMIE between AQ and Overall Gamma



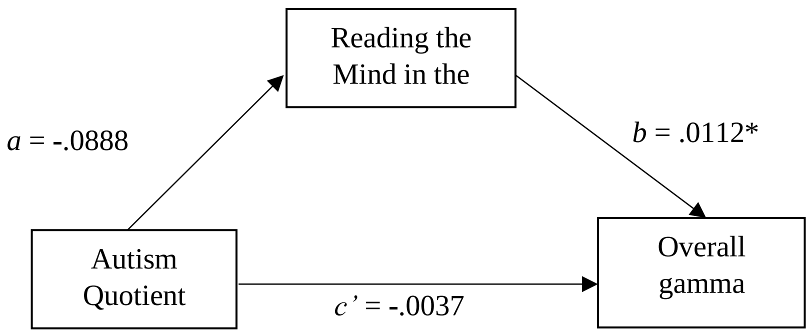
*p <.05, **p<.01, ***p<.001

The mediation analysis for AQ x *perceptual* gamma with RMIE as the mediator showed that the effect of AQ on overall gamma was not significant (B = -.005, SE = .004, t =-1.25, p = .21). The direct model was not significant (B = -.004, SE = .004, t = -.99, p = .32) and the indirect effect was not significant (B = -.001, SE <.001; (bootstrapped 95% confidence interval included zero, -.0031 to .0006), indicating that any relationship between AQ score and overall gamma was not mediated by RMIE.

.

Figure 14

Path Coefficients for the Mediating Effect of RMIE between AQ and Perceptual Gamma



*p <.05, **p<.01, ***p<.001

Experiment 1: Discussion

As predicted, the results from Experiment 1 showed that metacognitive accuracy was associated significantly with both mindreading (positive association) and autism traits (negative association), even when cognitive level performance was taken into account. Through examining metacognitive awareness for *semantic* and *perceptual* tasks separately it became apparent that both were associated with mindreading ability and remained so when we accounted for cognitive level performance. However, metacognitive accuracy was only associated with autism traits in the perceptual task, and this became marginal when cognitive level performance was taken into account. Despite these positive associations, mediation analysis showed that mindreading was not *mediating* the relationship between autism traits and metacognition (*overall* or *perceptual)*. This suggests that autism traits are related to metacognitive accuracy independently of mindreading.

In terms of metacognitive reaction time, there was a significant negative correlation between autism traits and metacognitive decision time on the *perceptual* task, but not the *semantic* task. In contrast to what was expected this implies that those with more autism traits were able to access metacognitive processing and interpret it quicker, and thus arrive at a wagering decision with relative ease. Although, this became marginal when object-level reaction time was controlled for. Furthermore, as highlighted above, this judgment was not necessarily more accurate given that autism traits were associated with lower accuracy. Unexpectedly, there was no significant correlation between mindreading and metacognitive reaction time on either the *perceptual* or *semantic* tasks.

In Experiment 2, we conducted a case-control study examining ~~of~~ these issues in a group of autistic adults and a group of neurotypical comparison adults. Based on the results from Experiment 1, it was predicted that the autism group would show significantly lower metacognitive accuracy than the typically developing group overall, and on both the *perceptual* and *semantic* tasks. It was also predicted that the autism group would be quicker at making their meta-level judgements overall and in the *perceptual* task. We predicted a significant association between metacognitive accuracy (overall, perceptual, and semantic) and mindreading, with better mindreading being associated with better metacognitive accuracy. It was also expected that there would be a significant association between metacognitive accuracy (*overall* and *perceptual*) and autism traits, with fewer autism trats being associated with better metacognitive accuracy. Finally, it was expected that autism traits would be associated with shorter meta-level reaction times overall and in the *perceptual* task.

Experiment 2: Method

Participants

Twenty-seven adults with a diagnosis of autism (14 males) and 27 typically developing (18 males) adults took part in the current study. Three participants (1 ASD, 2 TD) were excluded from all analysis due to a lack of variation in wagers placed in one or both metacognitive tasks, meaning it was not possible to calculate their gamma scores. This left 26 autistic participants (13 males) and 25 typically developing participants (13 males). All of the participants in the autism group had received a formal diagnosis in accordance with established criteria (American Psychiatric Association 2013; World Health Organization 1993). Out of the autistic participants we were able to obtain Autism Diagnostic Observation Schedule (ADOS) scores for 16 of them, nine whom scored seven or above. The mean ADOS total score among participants in the autism group was 7.38 (*SD* = 4.92), which is just above the clinical cut-off on this measure. The ADOS is a semi-structured observational assessment of social, imagination and communication skills and a score of ≥7 is consistent with a diagnosis of autism (Lord et al., 2000).

Details of participant characteristics can be seen in Table 20. Full Scale (FSIQ), Verbal (VIQ) and Performance (PIQ) IQ were assessed using the Wechsler Abbreviated Scale for Intelligence-II (Wechsler, 1999)[7]. All participants also completed the RMIE and an animation task as a measure of mindreading ability, and the AQ as a measure of autism traits. Participants also completed the BART as a measure of risk propensity. There were no significant differences between the autism and TD group in terms of age, FSIQ, VIQ, PIQ, or risk propensity. In line with their diagnostic status these was the expected between group differences in number of autism traits. Unexpectedly, there no significant difference in mindreading ability.

The study was ethically approved by the University's Research Ethics Committee (202116102756976964) and informed consent was obtained prior to commencing the tasks. All participants received payment of £10 per hour for their time, and all participants were provided with a written debrief following completion of all tasks, with the opportunity to contact the researchers should they have any further questions.

[7] Due to government restrictions imposed during the COVID-19 pandemic, 18 participants (9 ASD, 9 TD) completed three of the WASI subtests (vocabulary, matrix reasoning, and similarities) online. To obtain a full-scale IQ the block design subtest was given the same t-score as the matrix reasoning score.

Table 20

Experiment 2 Participant Characteristics: Means, Standard Deviations (in brackets), and Inferential Statistics for Autism and Typically Developing groups

	Group				
	ASD (n = 26)	TD (n = 25)	*t*	*p*	*Cohen's d*
Age	33.52 (10.24)	34.78 (11.31)	0.42	.68	0.12
Full-scale IQ	104.00 (12.22)	106.04 (11.94)	0.60	.55	0.17
Performance IQ	102.50 (17.56)	103.28 (13.73)	0.18	.86	0.03
Verbal IQ	104.58 (8.54)	107.36 (10.90)	1.02	.31	0.28
Autism Quotient	33.92 (8.04)	17.16 (7.30)	7.79	<.001	2.17
Reading the Mind in the Eyes	26.08 (3.97)	27.84 (3.68)	1.64	.11	0.46
Animations	5.46 (1.84)	6.21 (1.56) [a]	1.54	.13	0.45
Balloon Analogue Risk Task	23.65 (12.27)	26.97 (11.53)	1.00	.32	0.28

[a]based on scores for 24 typically developing participants due to one participant deciding not take part in the animation task.

Materials, procedure and scoring

Participants completed the AQ, RMIE, BART, and the perceptual and semantic wagering tasks, all of which are described above. The procedures for AQ, RMIE, BART and wagering tasks were the same as in Experiment 1, although the BART and wagering task involved earning money instead of points in Experiment 2. Participants also completed the animations task (Abell et al., 2000). This task has been widely used to assess mindreading abilities in both the general population and those diagnosed with autism (Wilson, 2021). During this task, participants are required to watch four short video clips of two triangles moving around. The clips are presented on a computer screen and, after watching each clip, participants are asked to describe what they think was happening in the video. Participants were allowed to watch each clip twice and responses are recorded using an audio recorder and later transcribed. Accurate responses required participants to attribute mental states, such as desire and intention, to the two triangles. Scores ranged from 0 to 2 for each clip, with higher scores indicating better mindreading abilities. Participants completed one practice trial prior to commencing the test trials.

Experiment 2: Results

Diagnostic Group Comparisons

Table 21 shows the means and standard deviations for each key variable. A 2 (Group: ASD/TD) × 2 (Condition: perceptual/semantic) mixed ANOVA indicated that there was significant main effect of condition, $F(1, 49) = 21.00$, $p < .001$, $\eta_p^2 = .30$, with participants getting a higher proportion correct in the semantic condition ($M = .74$, $SD = .02$) than the perceptual condition ($M = .67$, $SD = .01$). There was no main effect of group, $F(1, 49) =$

0.85, p = .36, η_p^2 = .02 and no significant interaction between group and condition F(1, 49) = 0.33, p = .57, η_p^2 = .007.

In terms of counters wagered, A 2 (Group: ASD/TD) × 2 (Condition: perceptual/semantic) mixed ANOVA showed a significant main effect of condition, F(1, 49) = 50.05, p = <.001, η_p^2 = .51, with participants placing higher wagers in the semantic condition (M = 3.72, SD = 0.08) than the perceptual condition (M = 2.95, SD = 0.12). There was a borderline significant main effect of group with a moderate effect, F(1, 49) = 3.40, p = .07, η_p^2 = .07. There was also no significant interaction between group and condition F(1, 49) = 0.12, p = .73, η_p^2 = .002.

The 2 (Group: ASD/TD) × 2 (Condition: perceptual/semantic) mixed ANOVA for gamma, showed that there was no significant main effect of condition, F(1, 49) = 0.59, p = .45, η_p^2 = .01, or group, F(1, 49) = 0.002, p = .96, η_p^2 <.001, and no significant group × condition interaction, F(1, 49) = 0.04, p = .84, η_p^2 <.001. This suggest that both groups performed to an equal level of metacognitive accuracy across both conditions. The 2 (Group: ASD/TD) × 2 (Condition: perceptual/semantic) mixed ANOVA for wagering reaction time showed that there was a significant main effect of condition, F(1, 49) = 96.66, p <.001, η_p^2 = .66, with reaction times being faster in the perceptual condition (M = 1.12, SD = .06) than the semantic condition (M = 1.62, SD = .05). There was no main effect of group, F(1, 49) = 0.41, p = .53, η_p^2 =.008, and no significant group x condition interaction, F(1, 49) = 1.95, p = .17, η_p^2 = .04.

Table 21

Experiment 2 Key Variables: Means and Standard Deviations (in brackets)

		ASD (n = 26)	TD (n = 25)
Overall	Object-level proportion correct	.72 (.07)	.70 (.09)
	Average Wager	3.17 (.67)	3.50 (.60)
	Gamma	.38 (.30)	.39 (.31)
	Meta-level Reaction Time	1.40 (.35)	1.34 (.34)
Perceptual	Object-level proportion correct	.68 (.07)	.65 (.11)
	Average Wager	2.80 (.88)	3.09 (.87)
	Gamma	.36 (.26)	.35 (.26)
	Meta-level Reaction Time	1.19 (.45)	1.05 (.37)
Semantic	Object-level proportion correct	.75 (.10)	.74 (.11)
	Average Wager	3.54 (.64)	3.91 (.55)
	Gamma	.41 (.49)	.42 (.54)
	Meta-level Reaction Time	1.61 (.34)	1.62 (.40)

Mindreading Unmatched Comparison

Given that there was no significant difference between the two groups in mindreading as measured by either RMIE or the animations task, and that mindreading ability is key to the theoretical debate, we created a group difference in mindreading ability and ran the metacognitive accuracy analysis again. First, we removed the participant that did not complete the Animations and then we removed the highest scoring autistic participants and lowest scoring typically developing participants on mindreading measures until there was a significant group difference in both RMIE, $t(44) = -2.52$, $p < .05$, $d = -0.74$, and Animations, $t(44) = -2.04$, $p = 05$, $d = -0.60$. This left 24 participants in the autism group and 22 participants in the typically developing group. Groups remained matched for baseline in age, FSIQ, VIQ, PIQ (all ps $\geq .33$, all $d \leq .29$) but differed on AQ, $t(44) = 7.40$, $p < .001$, $d = 2.19$, in line with diagnostic status.

Table 22 shows the means and standard deviations for each key variable. A 2 (Group: ASD/TD) × 2 (Condition: perceptual/semantic) mixed ANOVA for gamma showed that there was no main effect of condition, $F(1, 44) = .14$, $p = .71$, $\eta_p^2 = .003$, or group, $F(1, 44) = .006$, $p = .94$, $\eta_p^2 < .001$. There was no significant group × condition interaction, $F(1, 44) = 0.05$, $p = .83$, $\eta_p^2 = .001$.

The 2 (Group: ASD/TD) × 2 (Condition: perceptual/semantic) mixed ANOVA for meta-level reaction time showed that there was a significant main effect of condition, $F(1, 44) = 89.06$, $p < .001$, $\eta_p^2 = .67$. There was no main effect of group, $F(1, 44) = .29$, $p = .59$, $\eta_p^2 < .01$ and there was no significant group × condition interaction, $F(1, 44) = .83$, $p = .37$, $\eta_p^2 = .02$.

Table 22

Experiment 2 Key Variables: Means and Standard Deviations (in brackets) for groups unmatched on mindreading

		ASD (n = 24)	TD (n = 22)
Overall	Object-level proportion correct	.71 (.07)	.69 (.09)
	Average Wager	3.16 (.70)	3.45 (.62)
	Gamma	.37 (.31)	.38 (.32)
	Meta-level Reaction Time	1.39 (.36)	1.33 (.36)
Perceptual	Object-level proportion correct	.68 (.08)	.66 (.12)
	Average Wager	2.80 (.91)	3.07 (.92)
	Gamma	.35 (.25)	.37 (.26)
	Meta-level Reaction Time	1.17 (.46)	1.07 (.38)
Semantic	Object-level proportion correct	.74 (.10)	.74 (.11)
	Average Wager	3.51 (.66)	3.83 (.54)
	Gamma	.40 (.51)	.39 (.57)
	Meta-level Reaction Time	1.61 (.35)	1.60 (.40)

Correlations

To examine the relationship between autism traits (AQ), metacognition (gamma) and mindreading (RMIE) a series of correlational analyses was conducted. Given that there was no group difference in mindreading or metacognition, the analysis was conducted with the group data combined. All 26 ASD and 25 TD participants were included, given that the correlation analyses are essentially an individual differences approach. All results from the correlational analysis are presented in Table 23.

As can be seen in Table 23, RMIE was significantly associated with overall gamma, with higher RMIE scores being associated with higher gamma. This association remained significant when controlling for object-level performance, $r_s(48) = .27$, $p = <.05$. There was also a significant correlation between RMIE and *semantic* but not *perceptual* gamma, with higher RMIE scores being associated with higher semantic gamma. This association remained significant when controlling for object-level performance, $r_s(48) = .38$, $p = <.01$. There was a significant correlation between RMIE and *perceptual* but not *semantic* meta-level reaction time, with higher RMIE score being associated with longer wagering reaction times. This association reduced in strength and was borderline significant when controlling for object-level reaction time, r_s (48) = .22, $p = .06$.

According to Fisher's Z test there was a significant difference between the autism group and the typically developing group in the magnitude of the meta-level reaction time × AQ correlation on the perceptual task, z = 2.15, p <.05. This correlation was significant in the typically developing group, $r_s(23) = -.42$, $p <.05$, but not the autism group, $r_s(24) = .19$, $p = .18$. The meta-level reaction time × AQ correlation on the perceptual task remained significant when controlling for object-level reaction time, $r_s(22) = -.37$, $p = <.05$. There was also a marginally significant difference in the semantic gamma and RMIE correlations, z = -1.71, $p =.08$, with a significant association among the typically developing group, $r_s(23) = .57$, $p = <.001$, but not the autism group, $r_s(24) = .15$, $p = .23$. The semantic gamma × RMIE correlation for the typically developing group remained significant when controlling for object-level performance, $r_s = .58$, $p <.01$. Lastly, there was a marginally significant difference in meta-level reaction time and

RMIE correlations in the perceptual task, z = -1.73, p =.08, with a significant association among the typically developing group, $r_s(23) = .52$, p = <.01, but not the autism group, $r_s(24) = .06$, $p = .39$. The meta-level reaction time × RMIE correlation on the perceptual task remained significant when controlling for object-level reaction time, $r_s(22) = .57$, p <.05. All other Zs between meta-level performance (gamma and reaction time) and AQ and mindreading (RMIE and Animations) were < 1.57 and all ps > .12.

Further analysis revealed that there was no significant difference in the strength of correlations between RMIE and gamma for each condition, $Z = -.05$, $p = .32$. The RMIE × semantic gamma correlation increased in strength after controlling for object-level performance, $r_s(126) = .30$, p <.001. Likewise, the RMIE × perceptual gamma correlation remained significant after controlling for object-level performance, $r_s(126) = .18$, p <.05.

Table 23

Experiment 2: Correlations between Key Variables

Variables	2	3	4	5	6	7	8	9	10
1. Autism Quotient	-.24*	.06	.04	.07	-.03	.13	.09	.07	.13
2. Reading the Mind in the Eyes		.25*	.30*	.10	.36**	.11	.25*	-.01	.02
3. Animations			.00	.00	.01	-.12	-.12	-.15	.34*
4. Overall Gamma				.67***	.78***	.04	.10	.01	-.01
5. Perceptual Gamma					.16	-.13	-.11	-.11	.17
6. Semantic Gamma						.09	.20	01	-.15
7. Overall meta-level reaction time							.88***	.90***	-.17
8. Perceptual meta-level reaction time								.65***	-.22
9. Semantic meta-level reaction time									-.14
10. Balloon Analogue Risk Task									

Note. * *p*<.05, ***p*<.01, ****p*<.001,

Experiment 2: Discussion

The results from Experiment 2 replicated the findings from Experiment 1, showing that there was a significant association between mindreading ability and metacognitive accuracy, with better mindreading ability being associated with better metacognitive accuracy. This remained the case for metacognitive awareness in the *semantic* task but not in the *perceptual* task. This was surprising given that Experiment 1 found an association for both types of tasks. Although note, that the correlation in Experiment 2 was not statistically smaller in magnitude that Experiment 1, $Z = .60$, $p = .27$.

The results also showed that there was a significant association between autism traits and meta-level reaction time on the *perceptual* task but only in the typically developing group. This association was also found among the student sample in Experiment 1. Unexpectedly and in contrast to Experiment 1 there was a significant association between mindreading and meta-level reaction time for *perceptual* task, with better mindreading being associated with longer reaction times. Further analysis reveals that this association was only apparent in the typically developing group.

The group comparison results revealed that there was no significant between-group difference in metacognitive accuracy, or speed at which metacognitive judgements were made. Given that there was no significant difference in mindreading ability between the two groups this lack of difference in metacognitive performance is not entirely surprising. The one mechanism account only predicts differences in groups that have diminished mindreading ability and it appears that the autism group in this study were not impaired at mindreading. It is, however, surprising that there was no significant between-group difference in metacognitive accuracy, or speed at which metacognitive judgements were made, when a group difference in mindreading ability was created. This is somewhat paradoxical. Mindreading was related to metacognitive accuracy in both experiments in all groups, yet an autism group that performed poorly on a mindreading task nonetheless performed well on the metacognitive tasks.

General Discussion

This is the first study to examine metacognitive accuracy in relation to mindreading and autism on a semantic decision task using post-decision wagering. Moreover, it is the first study to examine metacognitive awareness of performance on a *perceptual* and *semantic* task in a common sample. The main finding from the current study is that there is a significant association between mindreading ability and metacognitive accuracy, with better mindreading ability being associated with better metacognitive accuracy. Importantly this association was found in Experiment 1 and replicated in Experiment 2, validating the findings. Moreover, the association held for the *semantic* task in both Experiments. Interestingly, the significant association between mindreading and metacognitive awareness for the *perceptual* task was only found in Experiment 1. This apparent association was not replicated in Experiment 2, although as previously noted the association in Experiment 2 was not significantly different from Experiment 1. This association contrasts with the findings from previous studies that have used post-decision wagering to measure accuracy of metacognitive awareness (Carpenter et al., 2019 – chapter 3).

The results also contrasted with Carpenter et al's (2019) findings when it came to the ease (speed) at which participants made their metacognitive judgements on the perceptual task. Carpenter et al., for example, found a significant association between mindreading and metacognition in both their student and typically developing samples, with better mindreading being associated with quicker judgement speed. Experiment 1 of the current study found no such association, and Experiment 2 found the opposite association among the typically developing sample (i.e., better mindreading = slower reaction time).

In terms of autism traits, there was a significant association between the number of traits and overall metacognitive accuracy in Experiment 1, with more autism traits being associated with poorer accuracy. When examined separately the association was only apparent for the *perceptual* task not the *semantic*. This finding replicates Carpenter et al.'s findings on their perceptual task. Surprisingly, these associations were not replicated in Experiment 2, although the association in Experiment 2 was not significantly different from Experiment 1, Z = -1.30, p = .10.

There was also a significant association between autism traits and meta-level reaction time for *perceptual* decisions, which was found among the student sample in Experiment 1 and the typically developing sample in Experiment 2. Unexpectedly, this implied that those with more autism traits were able to access metacognitive processing and interpret it quicker, and thus arrive at a wagering decision with relative ease. Although, as previously discussed, this judgment was not necessarily more accurate given that autism traits were associated with lower accuracy. Carpenter et al did not find any association in their experiments, although their results were of the same magnitude in their student sample but in the opposite direction.

The results from Experiment 2 also showed there was no significant between-group difference in metacognitive accuracy, or speed at which metacognitive judgements were made. This was surprising given the associations that were found in Experiment 1 and potentially challenges the one-mechanism account. However, as with the autistic group in chapter five, examination of the autism sample showed that there was no significant difference in mindreading. This is important given that the one-mechanism account only predicts difficulties in people with mindreading difficulties. The following results, however, may pose a challenge given that when a between group difference in mindreading was created there remained no significant difference in accuracy or speed. These findings are similar to the results from the studies presented in chapter three (Carpenter et al.) and four, in that it showed no between group difference in accuracy, however, they contrast with the findings from chapters three and four on speed at which decisions are made. Chapters three and four found that the autism group were slower at making metacognitive decisions despite equal levels of accuracy.

The findings from the current study are somewhat confusing given that on the one hand there are associations between mindreading ability and metacognitive awareness but on the other hand, when there is a group of participants with diminished mindreading ability, they do not show a deficit. One possible explanation is that the autism group are using an alternative strategy to achieve the same level of accuracy as the typically developing group. Indeed, this was what Carpenter et al. suggested, however, unlike their study the current study did not find a significant difference in meta-level judgement speed. Therefore, there is no evidence to suggest that autistic participants were using a different strategy, although there was no real grounds for testing this hypothesis in the current study and so such a theory cannot be ruled out. Further research tapping into the processes and patterns of meta-level judgments may well help to clarify this possibility. Overall, the lack of between group differences are in line with the predictions made by simulation theory and two-system theory but the associations are in keeping with the one system/theory-theory approach.

Another issue to address with the current study is the representativeness of the sample. Whilst all participants had received a diagnosis of autism from a trained clinician, we were only able to obtain ADOS scores for 67% of participants and just over half (56%) of these participants received a score indicative of autism (*M* = 7.38, *SD* = 4.92). This compares to 80% (*M* = 9.52, *SD* = 4.64) in Carpenter et al's study and 72% scoring over the cut off (*M* = 9.27, *SD* = 4.68) in chapter three. This may go some way to explaining the lack of between-group difference in meta-level reaction time.

Another key difference between the present study and Carpenter et al's study is that the present study was conducted online. There are various advantages and disadvantages with this type of administration, however specifically for this study, the social demands of online studies are greatly reduced. On the one hand this enables one to examine metacognition with much reduced social demand, on the other, in the real-world autistic people are having to make meta-cognitive judgments in daily life in far more socially demanding situations than online or laboratory settings. Therefore, if the social demand of the setting has an impact on metacognitive ability, then perhaps both lab and online experiments greatly underestimate the difficulties that autistic individuals have in performing meta-cognitive tasks day to day (Chevallier et al., 2014, Kenworthy, Yerys, Anthony, & Wallace, 2008; Ozonoff, 1995; White, Burgess, & Hill, 2009). Indeed, Maras et al's (2020) data provides a moderate difference (*d*=-.51) in overall gamma when comparing social (*M* = .36, *SD* = .69) versus online (*M* = .68, *SD* = .55) delivery of a metacognitive task. Although note, Maras et al's study did not examine online versus laboratory-based measures in the same participant group. This is perhaps an issue that requires more research.

Overall, it appears that whilst there is a relationship between mindreading and metacognition, autistic adults (with mindreading) are able to overcome this and perform to an equal level as typically developing adults. To confirm this, future research should take into account the overall pervasiveness of mindreading difficulties in autism and examine the extent to which this impacts on metacognition. It may be that whilst some autistic people are impaired in metacognition this may well depend on the pervasiveness and severity on mindreading difficulties.

Chapter Seven: Knowing What We Know: Using the Feeling-of-knowing and Remember-Know-Guess Paradigm to Examine Mindreading and Autism Spectrum Disorder in Relation to Metacognition in adults

Abstract

Background: The study presented in this chapter employs a feeling-of-knowing task and a remember-know-guess task to examine prospective and retrospective metacognitive accuracy in relation to autism and mindreading. To date only one study has used feeling-of-knowing to examine metacognitive accuracy in relation to mindreading among autistic adults, and no study has used a remember-know-guess task. ***Method*:** In Experiment 1, 124 students studied a list of word pairs (cue-target), after which they were presented with the cue word and asked to recall the target word. Immediately after each recall the participants made a feeling-of-knowing judgements as to whether they would recognise the word in future. After this they completed a recognition task of the incorrectly recalled items and then made a remember-know-guess judgement for each decision. For Experiment 2, 27 autistic and 27 typically developing adults completed the same task as that presented in Experiment 1. ***Results:*** Results from the student sample showed a significant positive association between metacognitive accuracy (as measured by feeling-of-knowing and remember-know-guess) and mindreading. There was also a significant positive association between autism traits and metacognitive accuracy as measured by feeling-of-knowing. Both these associations were replicated in Experiment 2 but only in the autistic sample. In line with all the studies presented in this thesis, there was no significant between group difference in meta-cognitive accuracy.

Metacognition, the ability to attribute metal states to oneself, enables people to predict and make sense of their own actions (Flavell, 1979; Nelson & Narens, 1990). It plays a key role in learning and decision making, and evidence shows that it contributes towards academic achievement independently of general intelligence (Flavell, 1979; Nelson & Narens, 1990; Veenman & Beishuizen, 2004). Metacognition has been linked to mindreading, the ability to attribute metal states to others. Many theorists, for example, argue that mindreading and metacognition rely on the same metarepresentational processes (Gopnik, 1990; Perner, 1991; Carruthers, 2009). Evidence supporting this "one-system" theory has been drawn from a variety of sources including, and perhaps most notably, autism research (Carruthers, 2009).

Autism is a neurodevelopmental disorder characterised by atypical social-communication, and restricted and repetitive behaviours. Evidence shows that autistic people also have significant difficulties with mindreading, which is also thought to underly the atypical social-communication in autism (Baron-Cohen et al. 2001, Brewer, Young & Barnett, 2017, Brunsdon & Happe, 2014, Happé, 1995, Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). It is perhaps not surprising then that one-system theorists have been drawn to this population to help clarify the processes underlying metacognition as well as mindreading. It has been claimed by proponents of the one-system account that, like mindreading, metacognition is also diminished in this group (Carruthers, 2009). Opponents of the one-system view, such as simulation theorists and two-system theorists, argue that metacognition is intact among autistic people (Goldman, 2006; Nichols & Stich, 2003). A critical issue with the autism research that these theorists have drawn upon is that it looks at *past* mental states rather than *current* mental states (Grainger, Williams & Lind, 2014). Simulation theorists and two-system theorists propose that it is only current mental states that are directly accessible without mindreading. Therefore, autism research needs to focus on methods that measure awareness of current mental states rather than prior mental states.

Current mental states can be measured using a variety of methods, the main ones being judgments-of-learning (JOL), feelings-of-knowing (FOK) and judgements-of-confidence (JOC). Each of these tasks require the participants to complete and cognitive/object-level task and to make a meta-level judgement regarding their object-level performance. For example, participants may be asked to rate how confident they are that their answer is correct (JOC), how likely it is that they have learnt a particular item (JOL) or if they think they will recognise a specific item (FOK). The stronger the association between the object-level judgement and the meta-level judgement the more accurate the metacognitive awareness. Research using such methods in autism is relatively sparse, with the majority of research employing judgments-of-confidence (See chapter two) and only two examining feeling-of-knowing (Grainger, Williams & Lind, 2014; Wojcik, Moulin & Souchay, 2013). Feeling-of-knowing tasks require participants to complete a task, such as answering some general knowledge questions, and then to make a feeling-of-knowing judgement on any items they get incorrect. Participants recognition for the correct answer is then tested. The association between the feeling-of-knowing judgement for each item and the recognition performance for each item indicates the level of metacognitive accuracy.

Wojcik, et al., (2013) examined feeling-of-knowing among 18 autistic children and 18 typically developing children across two different tasks. One task examined feeling-of-knowing accuracy on an episodic memory task and the other examined feeling-of-knowing accuracy on a semantic task. From this Wojcik and colleagues concluded that autistic children had diminished metacognition for episodic memory task, but not in the semantic memory. Grainger, et al., (2014) also examined feeling-of-knowing accuracy for episodic memory but in autistic *adults* and also found it to be diminished in comparison to typically developing adults. Grainger et al., also conducted some exploratory correlational analysis in light of the claim by one-system theorists that there should be a significant association between mindreading a metacognition. Contrary to the one-system predication, Grainger et al did not find a significant association between mindreading and feeling-of-knowing accuracy. Although, they acknowledge that their study did not have enough power to detect moderate associations and thus may have missed meaningful associations.

Given the limited amount of research into feeling-of-knowing accuracy in relation to autism and mindreading, the current study aimed to increase our knowledge base, particularly with respect to the potential association with mindreading. The current study also aimed to extend the existing literature by exploring another under-explored ability among autistic people, namely to distinguish between remembering and knowing. Remembering reflects autonoetic memory, meaning the memory is linked to time and space. Knowing on the other hand relates to noetic memory, meaning the memory is related to a feeling of familiarity and lacks the detail of autonoetic memories. Essentially, remembering is related to episodic memory whereas knowing is linked to semantic memory.

In remember-know tasks, participants complete a memory task and then indicate if they 'remembered' or just 'knew' the answer. For example, participants may be asked to learn a list of words and then later on they may be presented with the initial words mixed in with some previously unpresented words. The participants are then asked to identify if each word was presented in the original study list (old) or if it is a new previously unpresented word (new). Immediately after each 'old' judgment they are asked to state if they 'remembered' or just 'knew' the word. Participants are instructed to provide 'remember' responses if they can think of details concerning the time of encoding, (i.e., what they thought or saw). If they cannot recollect any details but recognise/recall the item because of a pure sense of familiarity, then they are asked to respond with 'know'.

Research has found that autistic people produce fewer 'remember' and more 'know' responses than typically developing people (Bowler, Gardiner & Gaigg, 2007; Bowler, Gardiner & Grice, 2000; Souchay et al 2012, Meyer, Gardiner & Bowler, 2014). This difference in responses has been interpreted as a differences in memory processing among autistic people in comparison to typically developing people. Thus, autistic people are said to rely on noetic memory more than typically developing people. The distinction between 'remember' and 'know' responses provides a dual-process interpretation of the remember-know paradigms, however, evidence suggests that remember-know responses can be an indicator of memory strength and are a kin to confidence in the memory being correct rather than just distinct processing (Dunn, 2004; Haaf et al., 2021; Rotello, Macmillan, Hicks, & Hautus, 2006; Wixted, 2007; Wixted, & Mickes, 2010). Thus, such paradigms can be used to assess metacognitive awareness. This paradigm has also been extended to include 'guess' responses given that there may be trials in which their memory does not reach the criterion for 'remember' or 'know' (Gardiner, Ramponi, & Richardson-Klavehn, 2002). Hence, 'remember' responses reflect stronger memory traces and more confidence in the memory than 'know' responses, and 'know' responses reflect stronger memory traces/more confidence than 'guess' responses. Given this interpretation, a strong association between responses and object-level performance (i.e., 'remember' for correct responses and 'guess' for incorrect responses) would indicate good metacognitive accuracy.

Combining feeling-of-knowing with a remember-know-guess (RKG) paradigm within a common sample enables us to examine the landscape of metacognition in autism by looking a different meta-level tasks within the same participant group. Given that feeling-of-knowing judgement are made prior to the completion of an object-level task and remember-know-guess judgements are made immediately after completing and object-level task, combining these paradigms also enables us to look at prospective and retrospective meta-level judgements within a common sample.

As highlighted in chapter five there is good reason to suspect a dissociation between prospective and retrospective meta-level judgments, with evidence coming from cognitive (Fleming, Massoni, Gajdos, & Vergnaud, 2016), neurological (Chua et al. 2009) and clinical research (Cosentino, 2014). Therefore, examining performance on prospective and retrospective judgements within a common sample as well as across a variety of meta-level task is crucial if we are to understand the extent to which metacognition is intact/impaired in autism. As with previous chapters, the study presented in this chapter adopts both a case control and individual differences approach to provide a comprehensive study of the association/dissociation between mindreading, autism and metacognition.

Experiment 1: Method

Participants

One hundred and forty students (19 male, and one participant who stated they were non-binary) from a university in the southeast of England took part in the experiment. Six participants were excluded from all analyses because it was not possible to calculate their gamma scores due to them obtaining 100% recognition. It was also not possible to calculate both gamma scores for a number of participants due a lack of variation in their metacognitive ratings (FoK = 7, RKG = 3). This left 134 participants in total, 127 with FOK gamma (18 male) and 131 (18 male) with RKG gamma. One hundred and twenty-four participants had gamma for both FOK and RKG (19 male). Each group contained 1 participant who stated they were non-binary. Key characteristics for each group are presented in Table 24. Participants received course credits in partial fulfilment of their degree. The study was ethically approved by the University's Research Ethics Committee (202116105416636965) and informed consent was obtained prior to commencing the tasks. All participants were given a written debrief following completion of the tasks and could contact the researchers for further debrief if required.

Table 24

Experiment 1 Participant Characteristics: Means and Standard Deviation (in brackets)

	FOK and RKG (n = 124)	FOK Only (n = 127)	RKG Only (n = 131)
Age	21.08 (6.17)	20.95 (5.98)	21.08 (6.24)
Autism Spectrum Quotient	19.77 (7.29)	19.61 (7.32)	19.77 (7.38)
Reading the Mind in the Eyes	24.76 (5.67)	25.06 (5.21)	24.69 (5.70)

Materials, procedure and scoring.

This was an online study implemented using Psychopy 3 (Peirce et al., 2019) and Qualtrics software (2020). Participants were instructed to use their own laptop or desktop in a quiet space without any distractions. Participants completed the Autism Spectrum Quotient as a measure of autism traits, and Reading the Mind in the Eyes, as a measure of mindreading ability. Participants also completed a feeling-of-knowing/Remember Know task, as a measure of metacognition. The stimuli were four to five letter nouns chosen from the MRC psycholinguistic database (Coltheart, 1981) all words within each trial were matched closely for Kucera-Francis written frequency (Kucera & Francis, 1967; See Appendix 3).

Feeling-of-Knowing/Remember-Know-Guess Task. This task consisted of three phases, *the study phase, the cued recall phase, and the recognition phase* (see Figure 15). Within the *cued recall* phase and *recognition* phase participants were required to make metacognitive judgements. Prior to completing the main task, participants completed a practice version involving five trials. The practice version did not involve a distractor task because it was purely to allow participants to familiarise themselves with the process, although they were made aware that there would be a task in-between the *study phase* and the *recall phase* for the main version.

Study Phase: During this phase, participants were presented with 50 word-pairs (cue and target) in the centre of a computer screen. Each word pair was presented on the screen for five seconds before moving on to the next word pair. Once all word pairs had been presented, participants complete a distractor task followed by a cued recall phase.

Cued Recall Phase: During this phase, participants were shown individually presented cue words and asked to recall the missing target word associated with each cue. Participants were then required to type the word directly into the computer. Immediately after each recall attempt, participants were asked to judge if they thought they would be able to recognize the missing target word when given a choice between four words. Participants could click on one of three boxes; No, Maybe, or Yes, to indicate their feeling-of-knowing judgement (see Figure 15). After making their feeling-of-knowing judgement, participants completed the recognition phase.

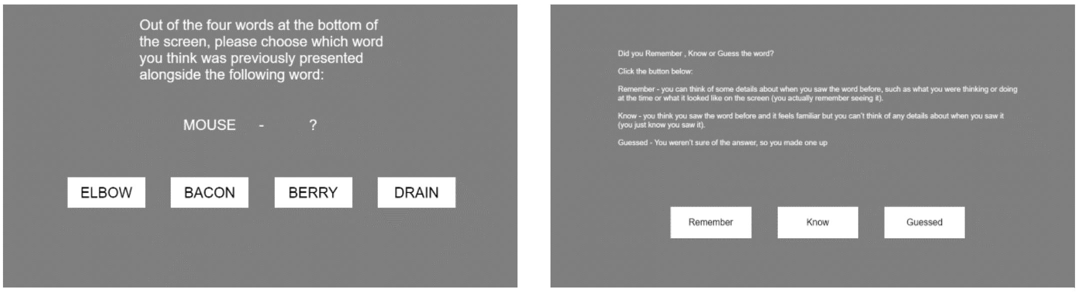
Recognition Phase: During this phase, participants were presented with a cue word in the centre of the screen with four other words directly below it (correct target, incorrect target, and two novel words). They were then asked to identify which of the four words were previously presented alongside the cue word. Immediately after this, participants were asked if they remember, knew, or guessed the target word. The definitions of remember, know and guess were presented at the top of the screen when making their decision. Figure 15 shows an example of a trial from this task.

Figure 15

Example Trial in the Feeling-of-Knowing/Remember-Know-Guess Task

Study Phase: The participant is presented with 50-word pairs (cue-target) and have five seconds to study each pair.

Recall Phase: The participant is presented with the cue word from each previously studied word pair and is asked to recall the correct target word. Immediately after entering the target word the participant makes a feeling-of-knowing judgement.



Recognition Phase: The participant is presented with the cue word and four potential target words, and must identify the correct target word. Immediately after making their choice the participant makes a remember-know-guess judgement.

From the feeling-of-knowing/remember-know-guess task two measures of metacognitive accuracy were calculated, ‘feeling-of-knowing accuracy’, and ‘remember-know-guess accuracy’. ‘Feeling-of knowing accuracy’ was indexed in each participant by calculating a gamma correlation (Kruskal & Goodman, 1954) between recognition accuracy and feeling-of-knowing response for incorrectly recalled items. To calculate feeling-of-knowing accuracy, ‘No’ responses were assigned a value of one, ‘Maybe’ responses were assigned a value of two, and ‘Yes’ responses assigned a value of three.

‘Remember-know-guess accuracy’ was also indexed by calculating gamma correlations between recognition accuracy and ‘remember-know-guess response’ for each participant. ‘Remember-know-guess responses’ were assigned a value of three, two, and one, respectively. Thus, large positive gamma scores indicate good metacognitive accuracy for both ‘feeling-of-knowing accuracy’, and ‘remember-know-guess accuracy’. For example, a large gamma between RKG responses and recognition memory indicates that the participants were more accurate in judging the strength of their memory. Metacognitive accuracy ranges from − 1 to + 1, with scores of 0 indicating chance level accuracy. This measure has been recommended by Nelson (1984), and Nelson et al. (2004) and has been extensively used in research on metacognitive monitoring processes (e.g. Grainger et al. 2016; Sawyer et al. 2014; Williams et al. 2018). Use of gamma in the current study also serves to facilitate comparisons with other studies of metacognition in autism, which have almost exclusively employed gamma as the main dependent variable.

Autism-Spectrum Quotient

The Autism-Spectrum Quotient is a widely-used and well-validated self-report measure of autism traits (AQ; Baron-Cohen et al. 2001a). It is considered to be a reliable measure of autism traits in both clinical and subclinical populations. The AQ presents participants with 50 individual statements (e.g., "I find social situations easy") and participants are asked to decide the extent to which they agreed with each statement by responding on a 4-point Likert scale, ranging from "definitely agree" to "definitely disagree". Higher scores indicate more ASD traits, with a maximum possible score of 50.

Reading the Mind in the Eyes Task

The Reading the Mind in the Eyes task is a widely used measure of mindreading among intellectually able adults, including those diagnosed with autism (RMIE; Baron-Cohen et al. 2001b). The task involves looking at photographs of eyes and deciding what the person in the picture is feeling. Participants are presented with 36 eye stimuli and are required to select an emotion that best described what the person in the picture may be feeling out of four possible emotions. Scores ranged from 0 to 36 with higher scores indicating better mindreading abilities.

Statistical Analysis

Where results are predicted a priori on the basis of a solid theoretical foundation and/or previous empirical findings, reported significance values are for one-tailed tests. All other significance values are two-tailed. All correlational analysis used Pearson's correlation except where the data was skewed in which case Spearman's rho was used. Where t-tests were used, we report Cohen's d values as measures of effect size ($\geq .0.20$ = small effect, ≥ 0.50 = moderate effect; ≥ 0.80 = large effect; Cohen, 1969). Where ANOVAs were used, we report partial eta squared (η_p^2) values as measures of effect size (≥ .01 = small effect, ≥ .06 = moderate effect, ≥ .14 = large effect; Cohen 1969).

Experiment 1: Results

The mean gamma score for FOK was .40 (*SD* = .52) and for RKG it was .66 (*SD* = .38), with both being significantly different from zero [FOK accuracy, $t(126) = 8.67$, $p < .001$, and RKG accuracy, $t(130) = 19.97$, $p < .001$]. This indicates that participants were significantly above chance in their judgment accuracy for both metacognitive judgements, meaning that participants were more likely to select 'yes' during the 'feeling-of-knowing judgement' and 'remember' during the 'remember-know-guess judgment' for words they correctly recognised. Furthermore, when we examined the participants that had scores for both FOK and RKG, there was a significant difference in meta-level performance between FOK judgements and RKG judgement, $t(123) = 5.19$, $p < .001$, $d = 0.48$, with participants performing more accurately in the RKG judgement ($M = .65$, $SD = .38$) compared to the FOK judgement ($M = .39$, $SD = .52$).

Correlations

There was a significant association between RMIE and both FOK gamma, $rs(127) = .21$, $p < .01$, and RKG gamma, $rs(131) = .22$, $p < .01$. The RMIE × FOK gamma, $rs(124) = .15$, $p < .05$, and the RMIE × RKG gamma, $rs(128) = .15$, $p < .05$, both remained significant when controlling for object-level performance. There was also a significant association between AQ and FOK gamma, $rs(127) = .16$, $p < .05$, but not AQ and RKG gamma $rs(131) = .06$, $p = .24$. The AQ × FOK gamma association remained significant and increased in size when controlling for object-level performance, $rs(124) = .18$, $p < .05$. The AQ × RKG gamma association, $rs(128) = .10$, $p = .13$, remained non-significant when object-level was controlled for but it was not significantly different to the AQ × FOK gamma, $z = .65$, $p = .26$.

Experiment 1: Discussion

As predicted, the results from Experiment 1 showed that there was a significant relationship between metacognitive accuracy and mindreading for both ‘feeling-of-knowing’ judgements and ‘remember-know-guess’ judgements. In both cases the association was positive meaning that the better an individual’s mindreading ability the more accurate their metacognitive judgements. Moreover, both associations remained significant when object-level performance was taken into account.

There was also a significant relationship between ‘feeling-of-knowing’ accuracy and autism traits. Surprisingly the association was in the opposite direction to what was predicted, with more autism traits being associated with increased accuracy. This association also remained significant when object-level performance was controlled for. In contrast, the correlation between autism traits and ‘remember-know-guess’ accuracy was non-significant.

For Experiment 2 it was predicted that there would be a significant association between mindreading and ‘feeling-of-knowing’ accuracy, and ‘remember-know-guess’ accuracy and mindreading, when object-level performance was controlled for. It was also predicted that there would be a significant association between number of autism traits an individual has and ‘feeling-of-knowing’ but not ‘remember-know-guess’ accuracy, with more autism traits being associated with better accuracy.

Considering the conflicting associations for the ‘feeling-of-knowing’ accuracy, (i.e., more autism traits associated with better metacognition and better mindreading associated with better accuracy) our group comparison predictions were based on previous research that found a significant between-group difference in metacognition when studying metacognitive accuracy in autistic adults. Therefore, it was predicted that the typically developing group would be significantly more accurate than the autism group in their feeling-of-knowing judgements, provided there was a significant between group difference in mindreading. It was also predicted that there would be a significant between-group difference in 'remember-know-guess' accuracy, with the typically developing group being more accurate than the autism group, again this was only expected on the basis that there is a between-group difference in mindreading. This is in line with the significant association between mindreading and 'remember-know-guess' accuracy.

Experiment 2: Method

Participants

Twenty-seven adults with a diagnosis of autism (14 males) and 27 typically developing (18 males) adults took part in the current study. Four autistic participants were excluded from all analysis because it was not possible to calculate either FOK gamma or RKG gamma scores due having 100% recall. A further seven autistic participants and three typically developing participants were excluded from the *group* analysis due to a lack of variation in either FOK responses, or RKG responses. This meant that it was not possible to calculate *both* gamma scores among these participants (only one of the two scores was available for analysis). This left 16 autistic participants (8 males) and 24 (15 male) typically developing participants in the group analysis.

All of the participants in the autism group had received a formal diagnosis in accordance with established criteria (American Psychiatric Association 2013; World Health Organization 1993). Out of these participants, we were able to obtain ADOS scores for 9 of them, four of whom scored seven or above (M = 8.00, SD = 6.00). The ADOS is a semi-structured observational assessment of social, imagination and communication skills and a score of ≥7 is consistent with a diagnosis of autism (Lord et al., 2000).

Details of participant characteristics for each group can be seen in Table 25. Full-scale (FSIQ), Verbal (VIQ) and Performance (PIQ) IQ were assessed using the Wechsler Abbreviated Scale for Intelligence-II (Wechsler 1999)[8]. All participants also completed the RMIE and the Animation task as measures of mindreading ability, and the AQ as a measure of autism traits. There were no significant differences between the autism and typically developing group in terms of age, FSIQ, VIQ, or PIQ. There were expected between group differences in number of autism traits (in line with their diagnostic status), but not mindreading ability, although the difference in Animation score was marginal and moderate.

The study was ethically approved by the University's Research Ethics Committee (202116102756976964) and informed consent was obtained prior to commencing the tasks. All participants received payment of £10 per hour for their time, and all participants were provided with a written debrief following completion of all tasks, with the opportunity to contact the researchers should they have any further questions.

[8] Due to government restrictions imposed during the COVID-19 pandemic, 17 participants (7 ASD, 10 TD) completed three of the WASI subtests (vocabulary, matrix reasoning, and similarities) online. To obtain a full-scale IQ the block design subtest was given the same t-score as the matrix reasoning score.

Table 25

Experiment 2 Participant Characteristics: Means, Standard Deviations (in brackets), and Inferential Statistics for Autism and Typically Developing groups

	Group				
	ASD (n = 16)	TD (n = 24)	*t*	*p*	*Cohen's d*
Age	34.87 (11.96)	34.44 (10.70)	0.12	.91	0.04
Full-scale IQ	99.25 (11.69)	105.50 (11.73)	1.43	.16	0.53
Performance IQ	96.13 (16.90)	103.13 (14.00)	1.66	.12	0.46
Verbal IQ	102.50 (8.41)	106.63 (11.05)	1.27	.21	0.41
Autism Quotient	33.06 (7.54)	17.50 (7.25)	6.58	<.001	2.11
Reading the Mind in the Eyes	25.75 (3.61)	27.54 (3.80)	1.49	.14	0.48
Animations	5.31 (1.78)	6.30 (1.51)[a]	1.87	.07	0.61

Note. [a]based on scores for 23 typically developing participants due to one participant deciding not take part in the animation task.

Materials, Procedure and Scoring

Participants completed the AQ, RMIE, and the feeling-of-knowing/remember-know-guess task, all of which are described above. In addition to these tasks, participants completed an Animations task (Abell et al. 2000). This task has been widely used to assess mindreading abilities in both the general population and those diagnosed with autism (Wilson, 2021). During this task, participants are required to watch four short video clips of two triangles moving around. The clips were presented online using video conferencing software (Zoom, 2021), where the researcher shared their screen and played each clip. After watching each clip, participants are asked to describe what they thought was happening in the video. Participants watched each clip twice and responses were recorded using the built-in recorder in Zoom and later transcribed. Accurate responses required participants to attribute mental states, such as desire and intention, to the two triangles, based on the criteria outlined in Abell et al. (2000). Scores ranged from 0 to 2 for each clip, with higher scores indicating better mindreading abilities. Participants completed one practice trial prior to commencing the test trials.

Experiment 2: Results

Diagnostic Group Comparisons

Table 26 shows the means and standard deviations for key variables for each group. There was a between group difference in the number of words correctly recalled, $t(38) = 2.20$, $p < .05$, $d = 0.66$. There was also a significant difference in the number of words correctly recognised, $t(38) = 2.05$, $p = 05$, $d = 0.63$.

In terms of remember-know-guess responses there was a significant difference in the proportion of ‘remember’ responses $F(38) = 5.67$, $p = .02$, $\eta_p^2 = .13$, with the typically developing group ($M = .37$, $SD = .27$) making more ‘remember’ responses than the autism group (*M* = .19, *SD* = .16). There was no significant between-group difference in the proportion of 'know', $F(38) = 0.20$, $p = .66$, $\eta_p^2 = .01$, or 'guess' responses, $F(38) = 2.80$, $p = .10$, $\eta_p^2 = .07$. There was also a significant difference in the proportion of 'yes' responses for the FOK judgment, $F(38) = 4.25$, $p = .05$, $\eta_p^2 = .10$, with the typically developing group providing more 'yes' responses (*M* = .13, *SD* = .13) than the autism group (*M* = .06, *SD* = .07).

A 2 (Group: ASD/TD) × 2 (Judgement type: FoK/RKG) mixed ANOVA was then conducted on gamma scores[9]. The analysis showed that there was a significant main effect of judgement type, $F(1, 38) = 4.26$, $p <.05$, $\eta_p^2 = .10$. There was no main effect of group, $F(1, 38) = 0.01$, $p = .93$, $\eta_p^2 <.001$, and no significant group × condition interaction, $F(1, 38) = 0.05$, $p = .83$, $\eta_p^2 <.001$.

[9] As with chapters five and six we 'unmatched' groups on mindreading and ran the analysis again. There remained no significant group difference in gamma, $F(1, 28)=0.40$,$p=.53$,η_p^2=01, and no group × judgement type interaction, $F(1, 28)=0.09$, $p=.72$, $\eta_p^2=.003$. The main effect of judgement type became marginal, $F(1, 28) = 3.32$, $p = .08$, $\eta_p^2 = .11$. There was no significant different in the proportion FOK responses, and the RKG responses reflected the same pattern of significant in the original analysis.

Table 26

Experiment 2 Key Variables: Means, Standard Deviations (in brackets), and Inferential Statistics

		Group	
		ASD (n = 16)	TD (n = 24)
Recall correct (proportion)		.13 (.17)	.28 (.26)
Recognition correct (proportion)		.44 (.21)	.60 (.28)
FOK	Average rating	1.74 (.41)	1.87 (.40)
	Gamma	.46 (.57)	.42 (.65)
RKG	Average rating	1.53 (.40)	1.87 (.54)
	Gamma	.65 (.33)	.66 (.54)

Correlations

To examine the relationship between autism traits (AQ), metacognition (gamma) and mindreading (RMIE) a series of correlational analyses were conducted. All participants that had at least one gamma score were included in the analysis[10]. This resulted in 20 autistic participants and 26 typically developing participants in the RKG analysis, and 19 autistic and 25 typically developing participants in the FOK analysis. Table 27 shows the results from the correlational analysis. There was a significant association between FOK gamma and AQ, this remained significant when the proportion of words correctly recalled was controlled for, *rs*(41) = .28, *p* <.05. All other associations were non-significant.

[10] Due to the high number of participants (Exp 1 = 17, 17, Exp 2 = 13) that would have been excluded in Exp 1 and 2 had we only maintained those that had both gammas, we chose to include all participants in the correlational analysis even if only one gamma score was available. We appreciate that this makes the analysis more complex and contrasts with previous chapters, however, the number of participants that this affects was not as high in the other studies (Chapter 4: Exp 1 = 0, Exp 2 = 4; Chapter 6: Exp 1 = 6, Exp 2 = 3).

Although note that the FOK gamma x RMIE association was not significantly different from Exp 1, z = 0.34, p = .37. The RKG gamma x RMIE association was also of a similar magnitude to that found in Exp 1, z = 0.53, p = .30.

Table 27

Experiment 2: Correlations between Key Variables

Variables	2	3	4	5
1. FoK Gamma	.27*	.26*	.15	-.05
2. RKG Gamma		-.04	.13	.12
3. Autism Quotient			-.25*	.18
4. Reading the Mind in the Eyes				.19
5. Animations				

Note. * p<.05,

Analyses showed there was a significant difference between groups for all correlations except for RMIE × RKG gamma. Table 28 shows the correlations for each group. There was a significant association between FOK gamma and both RMIE and AQ in the autism group but not the typically developing group. The AQ × FOK gamma became marginally significant once the proportion of words correctly recognised was controlled for, $r_s(16) = .39$, p =.06, but was marginally stronger than the same association in Exp 1 where it was significant, z = -1.34, p =.09. The RMIE × FOK Gamma was also not significant once the proportion of words correctly recognised was controlled for, $r_s(16) = .29$, p <.12. Although again it was not significantly different in magnitude to the same association in Exp 1 where the association was significant, z = -.86, p = .20. There was also a significant association between RKG gamma and both Animations and AQ in the autism group but not the typically developing group. The AQ × RKG gamma was not significant once the proportion of words correctly recognised was controlled for, $r_s(16) = .20$, $p = .21$, and the Animation x FOK gamma became marginally significant, $r_s(16) = .32$, $p = .09$.

Table 28

Experiment 2: Group Gamma x Key Variable Correlations

		Group		z	*p*
		ASD	TD		
FOK Gamma					
	Reading the mind in the Eyes	.43*	-.05	1.68	.05
	Animation	.26	-.28	1.83	.03
	Autism Spectrum Quotient	.59**	.14	1.77	.04
RKG Gamma					
	Reading the mind in the Eyes	.12	.12	<.001	.50
	Animation	.50*	-.17	2.38	<.001
	Autism Spectrum Quotient	$.37^{m}$	-.18	1.88	.03

Note. *$p < .05$, ** $p < .01$, *** $p < .001$, ^{m} <.06

Experiment 2: Discussion

There was no between-group difference in ‘feeling-of-knowing accuracy’ or ‘remember-know-guess accuracy’. There was a significant between-group difference in the proportion of remember responses, with the typically developing group providing more remember responses than the autism group. There was also a significant association between the *accuracy* of feeling-of-knowing judgements and autism traits, this remained significant when object-level performance was controlled for. This was in the same direction as that found in Experiment 1, with more autism traits being associated with better accuracy. Further inspection of the associations in each group revealed that the association between accuracy of feeling-of-knowing judgements and autism traits was only significant in the autism group. However, this association became only marginally significant once object-level was controlled for (though still similar in magnitude to the FoK × AQ partial correlation observed in Exp 1, which was significant in Exp. 1's larger sample).

There was also a significant association between mindreading (RMIE) and feeling-of-knowing judgement accuracy, with better mindreading being associated with better accuracy. This association also became non-significant once object-level was controlled for but again, it was not significantly different in magnitude from the association found in Exp 1 where it was significant. In terms of remember-know-guess accuracy there was a significant association with autism traits and mindreading (Animation) but neither were significant once object-level was controlled for.

General Discussion

This is the first study use a 'remember-know-guess' paradigm to assess metacognitive accuracy in relation to autism and mindreading. It is also only the second study to examine a prospective and retrospective online metacognitive accuracy in relation to mindreading and autism within a common sample using a common object-level task (see chapter five). The key finding from this study is that there is a significant association between mindreading and metacognitive accuracy as measured by feeling-of-knowing judgements and remember-know-guess judgments. In both cases better mindreading is associated with better accuracy. These association were found in Experiment 1 and, although non-significant in Experiment 2 where the sample size was considerably smaller, they were of a similar magnitude. This is in line with the findings from chapters five and six and the predictions made by proponents of the one-system theory (Gopnik, 1990; Perner, 1991; Carruthers, 2009).

In order to see if there were any differences in the associations between mindreading and metacognitive accuracy for autistic participants and typically developing participants each group was examined separately. This analysis revealed that the association between mindreading and feeling-of-knowing accuracy was significant in the autism group and, although not significant once the proportion of correctly recognised words was accounted for, it was still of a similar magnitude to that in Experiment 1. There was no association in the typically developing group, although as noted previously it was significant in Experiment 1 which included a much larger sample than the typically developing group in Experiment 2. In terms of the remember-know-guess accuracy and mindreading association for each group it remained non-significant, but both were similar in magnitude to that found in Experiment 1.

The analysis also revealed that there was a significant association between feeling-of-knowing accuracy and the number of autism traits. This was in the opposite direction to what was predicted by the one-system account and the results from experiment 1 presented in chapter five, with more autism traits being associated with better accuracy. This association was significant in Experiment 1 and 2. When the association was examined in each group the association was only significant in the autism group, although became marginal but not significantly different from Experiment 1 when the proportion of correctly recognised words was accounted for. The association in the typically developing group, whilst not significant was of a similar magnitude to Experiment 1. In terms of the association between remember-know-guess accuracy and the number of autism trait, the only significant association was found in the autism group, but this ceased to be significant once the proportion of correctly recognised words was accounted for.

The results also replicated the findings from previous research in terms of differences in remember-know responses, with autistic participants making significantly fewer ‘remember’ responses than the typically developing group (Bowler 2000, 2007; Souchay et al 2012; Meyer, Gardiner & Bowler, 2014). Moreover, the results from Experiment 2 revealed that the autistic participants were just as accurate as the typically developing participants for remember-know-guess judgments, suggesting that differences in remember-know-guess responses are the result of true difference in memory rather than a metacognitive issue.

There was also no significant between-group difference in the accuracy of feeling-of- knowing judgements, suggesting that autistic adults are just as accurate in such judgments as typically developing adults. This finding is in line with simulation theory and two-system theory, conflicting with previous findings and challenging the one-systems predictions (Grainger et al., 2014, Wojcik et al., 2013). These lack of between group differences also conflict with what would be expected considering the association between mindreading and metacognitive accuracy and the inherent difficulties with mindreading previously found among autistic adults.

Inspection of the performance on the mindreading tasks among both groups revealed that there was no significant between-group difference on any of the mindreading measures. This suggests that the autistic participants were just as good at mindreading as the typically developing group and thus did not have difficulties in mindreading. Therefore, from a one-system point of view this lack of difference in metacognitive accuracy is not surprising, one would not expect to find diminished metacognition in a group of participants with intact mindreading (Carruthers, 2009).

As discussed in previous chapters, there is evidence to suggest that the threshold to receive a diagnosis of autism has shifted, along with change in diagnostic criteria and a variety of other factors including increased awareness of autism (Mottron & Bzdok, 2020; Rødgaard, Jensen, Vergnes, Soulières, Mottron, 2019). It has been suggested that the combination of such factors has led to an increase in people receiving a diagnosis, with some even suggesting that autistic participants within case control studies are becoming less distinctive from typically developing populations (Mottron & Bzdok, 2020). Evidence of such a shift comes from the decreasing effect sizes for various cognitive constructs including mindreading (Rødgaard, Jensen, Vergnes, Soulières, Mottron, 2019). This may well explain the lack of between-group difference in mindreading among our groups, and potentially the lack of difference in metacognitive accuracy. However, when groups were unmatched on mindreading, as with previous chapters, there continued to be no between-group difference. Therefore, it seems that autistic adults as just as accurate in their meta-level judgments as typically developing adults despite mindreading difficulties. This challenges the one-system view and supports the predictions made simulation theorist and two-system theorists. However, the significant association, with more autism traits being associated with better metacognition, is of particular interest in the context of these findings. This association was in the opposite direction to what was predicted and is puzzling given the significant association between mindreading and metacognition, with better mindreading being associated with better metacognitive accuracy. Thus, this association combined with the lack of group difference may suggest that individuals with high autistic traits are able to achieve high levels of accuracy despite difficulties with mindreading (as was the case when groups were unmatched).

In keeping with previous chapters this points to the possibility that autistic adults may be able to use alternative strategies to reach the same level of accuracy as typically developing adults. In this case there maybe capacities linked to autism/autistic traits (but not mindreading) that enable autistic adults to overcome any difficulties with metacognitive accuracy or autistic individuals/individuals with more autism traits have had to rely on strategies that don't involve mindreading to achieve the same level of accuracy. Perhaps autistic adults draw on their metacognitive knowledge to make judgments about their object/cognitive-level performance. Certainly, there is evidence to suggest that metacognitive knowledge can be used to inform metacognitive decision and there is evidence to suggest that metacognitive knowledge is intact among autistic people (Farrant et al., 1999). Future research should investigate how metacognitive judgements can be made without metarepresentation (mindreading) and examine if autistic adults are using such processes.

Overall, our results replicate the results of previous studies, indicating that autistic adults are just as accurate as typically developing adults in their metacognitive judgements. It also leaves similar unanswered questions regarding the impact that the severity and pervasiveness of mindreading and autistic traits has on metacognitive accuracy. In addition, it opens up research into how autistic people are able to achieve good levels of metacognitive accuracy despite its link with mindreading difficulties. Considering the impact that metacognition has on daily functioning answering these questions will be fundamental to informing future research and practice.

Chapter Eight: General Discussion

The purpose of this thesis was twofold, first it aimed to establish the extent to which metacognition is intact among autistic people. The second objective was to clarify the theoretical debate regarding the relationship between mindreading and metacognition. Specifically, do mindreading and metacognition rely on the same metarepresentational processes? To address these questions, the thesis presented multiple studies using novel methods to examine the relationship between mindreading, metacognition and autism.

The focus was on the awareness of *current* mental states, rather than past metal states, because it is this awareness that lies at the heart of the theoretical debate. Until recently, research had focused on the awareness of past mental states, which do not necessarily require meta-representational processes. The initial meta-analysis and critical review in chapter two amalgamated the previous research examining online metacognition in autism. Through this it became clear that only 17 studies have examined this topic. The initial analysis indicated that metacognition is diminished among autistic people overall, though further scrutiny of the research revealed several key factors fundamental to our understanding of metacognition in autism. Firstly, the type of meta-level judgement being examined appears to have an impact on the level- of accuracy, with judgements-of-confidence and feelings-of-knowing being diminished but judgements-of-learning being apparently intact. This suggests that metacognition may not be globally impaired among autistic people. On closer inspection it also became apparent that it may be important to distinguish between children and adults. Children, for example, showed a moderate impairment whereas adults showed a relatively small impairment when compared to their typically developing counterparts. Thus, the age group of participants is an essential factor to consider when examining metacognition in autism. Another key issue was the considerable variation in the object-level task employed by each study, making it difficult to draw firm conclusions regarding the accuracy of meta-level judgements for different object-level tasks. This is important considering that meta-level judgements about different object-level tasks employ additional processes depending on the object-level task (Baird et al., 2013; Rouault, McWilliams, Allen & Fleming, 2018).

The fragmentation of research highlighted the need for further research, paying particular attention to the type of meta-level judgements being made, as well and the type of object-level task employed. Considering that only six out of the 17 studies had examined metacognition in autistic adults, along with the considerable variation in methods and outcomes, this thesis focused on metacognition in autistic adults[11]. In line with the recommendations from the meta-analysis, the thesis examined the landscape of strengths and weaknesses across different meta-level judgements and object-level tasks.

Out of the six adult studies included in the meta-analysis, four used judgements-of-confidence as a measure of metacognition. Judgements-of-confidence require participants to explicitly state their meta-level judgements by rating how confident they are in their object-level performance. However, many of our day-to-day meta-level judgements do not necessarily involve explicit verbal responses, and therefore research employing behavioural measures of metacognition are required to extend our understanding of metacognition in autism. Thus, chapter three and six employed a post-decision wagering paradigm, which is considered be the behavioral equivalent of judgements-of-confidence (Dienes & Seth, 2010; Ruffman, Garnham, Import, & Connolly, 2001; Persaud et al., 2007). Chapter three also examined the ease at which meta-level judgements are made using meta-level reaction times, this is an issue never explored before and goes beyond pure performance.

[11] The project intended to examine metacognition among autistic children as well as adults, however, due to government restrictions during the COVID-19 pandemic it was not possible to complete these studies.

Considering that the majority of studies into metacognition and autism examined judgements-of-confidence accuracy, but have never examined *ease* of meta-level judgements, chapter four took up this issue and examined judgments-of-confidence *accuracy* and *reaction time*. Chapter five and seven addressed the issue of comparing meta-level judgements by examining prospective and retrospective meta-level judgements within a common sample, and chapter six tapped into variations in object-level tasks and examined metacognitive awareness using a post-decision wagering paradigm for perceptual and semantic discriminations within a common sample.

Diagnostic comparisons

It is clear from the evidence presented in this thesis that autistic adults are just as accurate in their meta-level judgements as typically developing adults, with no between group differences in meta-level performance across any of the studies. Thus, autistic adults are able to make meta-level judgements to an equal level of accuracy as typically developing, age and IQ matched comparison participants. These results conflict with previous research that has found metacognitive accuracy to be diminished in autistic adults (Cooper et al., 2016; Nicholson et al., 2019; Grainger et al., 2014), but are in keeping with research such as by Maras et al. (2020) and Sawyer et al.'s (2014), which found metacognition to be intact. From a theoretical perspective, these results appear to contrast with predictions made by one-system theorists (Carruthers, 2009) and are in keeping with the predictions made by simulation (Goldman, 2006) and two-system theorists (Nichols & Stich, 2003). Thus, the evidence presented in this thesis regarding between-group differences in metacognitive accuracy support the simulation theory and the two-system theory of mindreading and metacognition. However, see discussion below regarding lack of between group differences in mindreading ability in three out of the five studies presented in this thesis (Table 31, p 250).

The evidence regarding the speed with which meta-level decisions are made was less straightforward. Two of the four tasks that assessed the speed at which meta-level judgments are made (chapters three & four) and indicated slower meta-level decision making among autistic adults than among typically developing adults. Importantly the difference in meta-level reaction time persisted even when object-level reaction time was taken into account. This indicates that the difference in speed found in chapters three and four was specific to *meta*cognitive, rather than cognitive, processes. Moreover, this difference was found across post-decision wagering and judgments-of-confidence, suggesting that the difference persisted across different metacognitive tasks/task modalities. As previously discussed, the speed at which meta-level judgments are made is considered to reflect the effort/ease with which meta-level decisions are made, just as when object-level decisions are made.

Interestingly, out of the two tasks that did not find a difference in meta-level reaction time, one was identical to the that used in chapter three, where there *was* a difference in meta-level reaction time. There are several possible explanations for the conflicting between-group results. First, both the tasks in chapter six were conducted online, whereas the tasks reported in chapters three and four were conducted in a laboratory setting. There are several potential issues with this, first, online studies do not enable control of external confounds such as the environment in which the tasks are completed. Hence, one cannot be sure the of level of distractions at the time of task completion and thus the level of attention participants paid to completing the tasks.

The second possible difference between the online version and the lab-based version is that there may have been variation in how accurate/sensitive the software was. However, evidence concerning reaction time data for online studies indicates that this is this is unlikely to explain the contrasting findings (Bridges, Pitiot, MacAskill, & Peirce, 2020). A large-scale study by Bridges et al, for example, examined a variety of software packages utilising a mix of web browsers. They found that the software used for online studies is accurate to at least 10ms. More specifically, they found that the software used in our study (Psychopy) was accurate to 3.5ms. Considering that the difference between our autistic and typically developing group was 260ms in chapter three and 300ms in chapter four, the software used for our online studies would have been precise enough to detect differences. Therefore, this is unlikely to explain the conflicting meta-level reaction time results.

Another possible explanation is the level of social demand placed on the participants in each version of the tasks. In the laboratory study there would have been a higher level of social demand in comparison to the online study. Some have suggested that when social demand is reduced, autistic participants can perform just as well on cognitive tasks as typically developing participants (Chevallier et al., 2014, Kenworthy, Yerys, Anthony, & Wallace, 2008; Ozonoff, 1995; White, 2013; White, Burgess, & Hill, 2009). However, others have questioned this, demonstrating that no difference is found between computer-based and experimenter-administered tasks when measuring higher-order cognition among autistic participants in a common sample (see Williams & Jarrold, 2013). In the context of the studies presented in this thesis, if the social demand of the laboratory version was responsible for the between-group differences in reaction time then one would also expect to see impairments across all other variables, including object-level performance. Our results show that this was not the case for either of the studies presented in chapters three or four. Thus, it seems highly unlikely that the additional social demand of the laboratory version tasks was responsible for the between-group difference in meta-level reaction time.

A more plausible explanation for the contrasting findings in meta-level reaction time is the differences in participant characteristics of the autism groups. In the study where no difference was found (chapter six) we were only able to obtain Autism Diagnostic Observation Schedule (ADOS) scores for 67% of participants[12]. Therefore, we cannot be sure of the symptom severity of the autistic participants in this study. It may be that the autistic adults that took part in the study presented in chapter six had milder symptoms than the groups in chapters three and four. Indeed, out of the participants that had ADOS scores, only 56% scored above the cut off indicative of autism. This compares to 70% in chapter three and 67% in chapter four. Thus, differences in the severity of symptoms, as measured by the ADOS, may well explain the contrasting findings. Certainly, this is an area of research that requires more investigation. Future research could, for example, examine metacognitive accuracy in relation to symptom severity using a variety of measures including the ADOS. This would inform researchers and clinicians at which point metacognition is likely to become an issue if at all.

Correlations

Throughout this thesis a series of bivariate correlations were conducted to examine the relationships between metacognition, mindreading and autistic traits. Partial correlations were also conducted, controlling for object-level performance. The reason for this is that object-level performance can have an impact on metacognitive accuracy. Indeed, performance on a cognitive-level task often correlates with metacognitive accuracy (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Thus, the easier the task is, the easier it is to detect one's own mistake. Therefore, bivariate correlations can make type 1 errors more likely (falsely rejecting the null). Conversely, controlling for object-level performance is a conservative/stringent

[12] We had intended to administer the ADOS for all participants, however, due to government restrictions we were unable to invite participants to the laboratory to conduct the assessment.

approach, given that cognitive and metacognitive levels influence each other in a bidirectional manner (Nelson & Narens, 1990). Therefore, controlling for object-level performance will mask some of the independent relationship between metacognitive accuracy and mindreading, increasing the risk of type 2 errors (falsely accepting the null). This is a particular issue for studies with small sample sizes due to reduced power to detect true underlying effects. Thus, the following conclusions will be based on partial correlations, controlling for object-level performance with the caveat that this may mean some meaningful results could missed, particularly where sample sizes are small such as in chapters three and four, and Experiment 2 of chapters five, six and seven.

Turning our attention to the results of the correlational analysis, Table 29 presents the association between key variables across all five studies presented in this thesis. As can be seen in Table 29, the majority (11/16) of tasks showed a significant association between *mindreading and metacognitive accuracy*. All of which indicated that poorer mindreading ability was associated with less accurate meta-level judgments. This is in keeping with the one-system view. However, once object-level was controlled for six of these remained significant, and two became marginally significant. Thus, accepting the more stringent partial correlations provides inconclusive evidence either way, with half being significant or close to significance and half showing no significant relationship between mindreading and metacognitive accuracy.

Table 29

Overview of Statistically Significant Results

	Chapter Three (PDW)		Chapter Four (JoC)		Chapter Five (Pro/Retro JOC)				Chapter Six (Perceptual/Semantic PDW)				Chapter Seven (FOK/RKG)			
	Exp 1	Exp 2	Exp 1	Exp 2	Exp 1		Exp 2		Exp 1		Exp 2		Exp 1		Exp 2	
					Pro	Retro	Pro	Retro	Per	Sem	Per	Sem	FOK	RKG	FOK	RKG
Gamma	-	No	-	No	-	-	No[a]	No[a]	-	-	No	No	-	-	No	No
Reaction time	-	Yes*	-	Yes*	-	-	-	-	-	-	No	No	-	-	-	-
Mindreading x Gamma	No	No	No	No	Yes[1*]	Yes[1m]	Yes[1,2]	Yes[2]	Yes[1*]	Yes[1*]	No	TD[1*]	Yes[1*]	Yes[1*]	ASD[1]	ASD[2m]
Mindreading x RT	Yes[1*]	TD[3*]	Yes[1*2*]	TD[1*,3*]	-	-	-	-	No	No	TD1[*o]	No	-	-	-	-
Autism traits x Gamma	Yes*	TD[*]	No	No	Yes*	Yes	No	TD	Yes*	No	No	No	Yes[o]*	No	ASD[om]	ASD[mo]
Autism traits x RT	No	No	No	Yes*	-	-	-	-	Yes[o]	No	TD[o*]	No	-	-	-	-

[1] = RMIE, [2] = Animation, [3] = MASC, TD = significant in typically developing group only, ASD = significant in autism group only

*= remained significant (p<.05) when object-level performance/ reaction time controlled

[a =] Marginal when 'unmatched' on mindreading, [m] = marginal

[O] = Opposite direction to expected

In terms of the relationship between *mindreading and the speed* at which meta-level judgments were made, five out of the eight tasks that measured reaction time found a significant association. Four of which indicated that better mindreading ability was associated with quicker meta-level reaction times, with one showing the opposite. These all remained significant once object-level reaction time was controlled for. Thus, these results are mixed given that only half found a significant association between mindreading and meta-level speed.

In terms of autism traits, there was a significant association with metacognitive accuracy in nine out of the 16 tasks. Six of the nine indicated that the more autism traits an individual has the poorer the metacognitive accuracy, and three of which indicated the reverse with more autism traits being linked to better accuracy. Of the six that showed a negative correlation, four remained significant when object-level performance was controlled for. Of the three that showed a positive correlation, only one remained significant once object-level was controlled for, the other two became marginally significant.

Finally, three out of the eight tasks revealed a significant association between autism traits and meta-level reaction time. Two out of the three indicated quicker reaction time for individuals with more autism traits, one of which remained significant once object-level reaction time was controlled for. The remaining significant association showed the reverse, more autism traits being associated with longer meta-level reaction times, this remained significant once object-level was controlled for.

Overall, the results from the correlational analysis provide mixed results, making it difficult to draw any firm conclusions. Throughout the thesis there is evidence of associations between mindreading and metacognition, both in terms of accuracy (see chapters five six and seven) and ease (see chapters three and four). There is also evidence of an association between autism traits and metacognitive accuracy (see chapters three, five and six). However, once object-level is controlled for the number of significant associations decrease. Therefore, further research is required to make sense of these inconsistent findings. Furthermore, there are several issues that must be addressed regarding the findings in this thesis in terms of both associations and group differences. The next section will examine how the differences in meta-level reaction time can be interpreted, followed by a review of the measures used throughout this thesis, and a discussion of some more general issues associated with the studies presented in this thesis as well as autism research more generally.

Why do autistic adults take longer to make meta-level decisions?

Does fast really mean easy?

Throughout this thesis it has been argued that the speed with which meta-level decisions are made equate to the ease with which such judgements are made. There is, however, an alternative explanation for the increased meta-level speed other than the ease at which participants made their meta-level decisions. It could be that autistic participants are merely more cautious or have a higher threshold before committing to a decision. The drift diffusion model, proposed by Ratcliff and Starns (2009, 2013), posits that individuals accumulate information before making their confidence judgment and that each decision has a boundary, thus once that threshold has been met the participant will make their decision. Therefore, it could be argued that autistic participants meta-level reaction times are longer because they have a higher threshold for the amount of evidence needed in order to commit to a meta-level decision. The current thesis did not measure this possibility and thus research is required to see if autistic participants would have been as accurate given the same amount of time as the typically developing participants. Future research, for example, could restrict the time autistic participants have to make their meta-level judgements to that of a typically developing population (< 2 seconds). This would enable us to see if the autistic participants are able to make equally accurate metacognitive judgements in the same time frame. If so, this may then indicate that autistic participants are merely being more cautious before committing to a decision rather than finding it harder.

Are Autistic Adults Compensating?

It is interesting that whilst autistic adults are just as accurate and typically developing, age and IQ matched comparisons, they appear to be slower to make such meta-level judgements. It is also curious as to why autistic children are impaired but autistic adults are not (see chapter two). It may be that any childhood deficit has resolved by adulthood, however, one other possible explanation that must be addressed is that autistic adults are using an alternative strategy, that takes longer, to perform well on meta-level tasks.

The notion that autistic adults may be using alternative strategies to perform well on tasks despite atypical conceptual competence is not a new concept and has been found across a variety of domains (Bowler, 1992; Hermelin & O'Connor, 1985). Indeed, this was precisely what Williams and Happé (2009) concluded when they examined metacognitive awareness as measured by a modified unexpected contents task. In 'unexpected contents' tasks participants are shown that a box that contains something different to what is labelled on the box. Traditionally, participants are then asked to verbalise what they believe to be in the box before they are shown its true contents. They are then asked a ‘self-question’, tapping into their prior false belief about what was in a box. They are also asked the ‘other-question’, which taps into what someone else will (falsely) believe is in the box. Research using this version finds that autistic children perform just as well on the self as they do on the other question. In Williams and Happé’s version, however, children were not required to verbalise their own false belief before being asked the self-test question. The results from their study showed that autistic children found the self-question significantly more difficult than the other-question. Crucially, non-autistic children performed consistently across both questions as they do when they are given the traditional version. Thus, Williams and Happé concluded that autistic children perform consistently on the self and other questions in the traditional version because they rely on recalling their previous statement to succeed rather than the metarepresentational resources that are used by non-autistic participants.

Thus, the concept that autistic adults use atypical strategies to achieve the same level of metacognitive accuracy is a perfectly plausible explanation for the pattern of results found across the studies presented in this thesis. This explanation also fits well with evidence that adults with autism tend to rely on deliberative reasoning strategies to solve cognitive tasks, rather than relying on intuitive processes employed by typically developing adults (Ashwin & Brosnan, 2019; Brosnan, Ashwin, & Lewton, 2017; Brosnan, Lewton, & Ashwin, 2016). According to Dual-Process theory (Evans and Frankish 2009), human decision-making is underpinned by two forms of reasoning. Reasoning based on heuristics (non-analytic), which tend to be fast, easy and intuitive (Type 1), and reasoning based on analytic processes, which tend to be slower, more effortful, and deliberative (Type 2). This notion fits well with the overall findings from this thesis and may also explain previous findings in the literature. Whereas typically developing adults from the general population tend to employ type 1 reasoning when completing metacognitive monitoring tasks, adults with autism tend to employ type 2 reasoning, which results in similar levels of accuracy but after a longer period of processing. Intellectually able autistic adults have already been through an education system that encourages the development of metacognitive skills, so arguably type 2 reasoning about mental states becomes ingrained as a response to training and difficulties with intuitive monitoring earlier in life.

Future research should therefore examine the processes underlying autistic adults' metacognitive decisions. For example, research could examine performance on tasks that tap into heuristic/type 1 based processing such as 'feeling of rightness' or even manipulate the information available and/or fluency/familiarity of processing when making meta-level judgments to see the extent to which such manipulations affect accuracy in autistic participants compared to typically developing participants. If differences are found, then this would indeed add further support to the theory that autistic adults use different processes to achieve the same level of accuracy as typically developing adults.

Mindreading Tasks

Three mindreading measures have been employed within this thesis, the Animations task, Reading the Mind in the Eyes (RMIE), and the Movie for the Assessment of Social Cognition (MASC). Whilst it has been assumed that these are all assessing the same construct (mindreading), it is possible that they are tapping into distinct aspects of mindreading or indeed different processes. Certainly, they all vary in the type of mental state the respondent is expected to attribute and the cues available for attributing those mental states. RMIE, for example, focuses purely on attributing emotions from photos of eyes, whereas the Animation tasks requires participants to attribute intentions, desires and beliefs from video clips of triangles moving around, and the MASC focuses on attributing thoughts, emotions and intentions from a video or a naturalistic social interaction between a group of people.

RMIE is the most commonly used measure of mindreading and has been utilised thousands of times, with good test-retest reliability as well as being quick and easy to administer (Fernández-Abascal, Cabello, Fernández-Berrocal, & Baron-Cohen, 2013). It also correlates with other measures of mindreading (Jones et al., 2018) and activates the same brain regions as other mindreading measures (Schurz et al., 2014). Furthermore, it places less demand on the participant to integrate situational details, allowing participants to focus purely on metal state attribution (Chung, Barch & Strube, 2011). Despite these advantages, some argue that RMIE is actually a measure of emotion recognition rather than mindreading (Oakley, Brewer, Bird & Catmur, 2016). The task involves participants looking at photographs of eyes and then selecting what the person may be feeling out of four possible emotions. It is argued that success on this task may be related to issues with emotion recognition rather than theory of mind. This has been supported by research showing that Alexithymia (a trait characterised by poor recognition of one's own and others' emotions) is a predictor of performance on RMIE (Oakley, et al., 2016). However, Nicholson et al's (2018) study showed no difference in performance on RMIE between autistic participants with and without alexithymia.

Considering the criticisms of RMIE, two other measures of mindreading were used within this thesis, the MASC and the Animation task. The MASC presents participants with naturalistic videos of people engaging with each other and it requires participants to infer thoughts, emotions, intentions. It also includes interpretation of pragmatic language such as sarcasm, faux pas, metaphors, as well as false belief and deception. This clearly extends attribution of mental states beyond emotions, however, success on the MASC also involves additional skills such as executive functions, central coherence, and receptive and pragmatic language thus it is not a measure of "pure" mindreading (Baron-Cohen et al., 1997; Heavey et al., 2000; Roeyers et al., 2001). Indeed, in the studies that used this measure within this thesis the autistic participants performed poorly in comparison to typically developing participants on the control questions as well as the mindreading questions (though not to as great an extent as on the mindreading questions).

Finally, the Animation task was also employed across the studies presented within this thesis. In this task participants watch silent video clips of triangles and are required to describe what is happening in the clip. Success on the task requires participants to attribute desires, beliefs and intentions to the triangles. As with RMIE, evidence shows that when participants complete the Animations task the same brain regions are activated as when they complete other mindreading tasks (Castelli et al., 2002; Castelli, Happe, Frith & Frith, 2000; Perner, Aichhorn, Kronbichler, Staffen & Ladurner, 2006; Samson, Apperly, Chiavarino,& Humphry, 2004; Saxe & Powell, 2006). Like the MASC, the Animations task requires participants to attribute metal states beyond just emotions and thus overcomes the main criticism of RMIE. It also has an advantage over the MASC in that does not rely as heavily on executive control and removes any reliance on receptive and pragmatic language ability. It does, however, still rely on expressive verbal ability, thus as with the MASC it is not a measure of "pure" mindreading. This is a limitation across all mindreading measures and is not unique to the measures used within this thesis. Thus, overall, there appears to be no single measure that captures the complexity of mindreading sufficiently and purely, therefore it is important to use a variety of measures of mindreading considering that it is a multifaceted construct and that different tools assess different aspects of same construct.

Metacognitive Tasks

There are various meta-level judgements that can be used to measure metacognition. This thesis chose to focus on four main measures, judgements-of-confidence, post-decision wagering, feeling-of-knowing, and remember-know-guess. Each of these methods require participants to complete a cognitive/object-level task and make a meta-level judgement regarding performance on that task.

Confidence and post-decision wagering

Chapters four and five of this thesis employed judgements-of-confidence, one of the most commonly used measures of metacognition (Dunlosky & Metcalfe, 2009). As previously described, the judgement-of-confidence paradigm requires participants to complete a cognitive/object-level task, and then rate how confident they are that their response was correct. In chapter five one of the tasks required participants to make their confidence judgment before completing the cognitive-level task. Thus, rather than tapping into retrospective metacognitive accuracy, as the classic judgement-of-confidence paradigm does, one of the tasks presented in chapter five tapped into prospective metacognition. Regardless of when the confidence rating is made, judgments-of-confidence require participants to make a judgment about their cognitive performance, and thus meta-represent their cognitive performance and reveal some level of insight into their cognitive-level performance. The extent to which participants' confidence in the accuracy of their response corresponds to the actual accuracy of their response indicates their metacognitive accuracy. As discussed previously this is often done by calculating a gamma correlation (Kruskal & Goodman 1954), with scores ranging from − 1 to + 1, where scores of 0 indicate chance level accuracy, and large positive scores indicate good metacognitive accuracy. Evidence indicates that confidence ratings provide a good measure of conscious awareness, or awareness of what one knows (Dienes & Seth, 2010). It does, however, rely on verbal ability and subjective reporting, and even with good verbal skills, the term ‘confidence’ is open to much interpretation and so people may use different criteria for each confidence level (Sandberg et al., 2010). It is also susceptible to metacognitive bias, the tendency to give high or low confidence ratings (Fleming & Lau, 2014). Thus, alternative, more concrete behavioural measures of metacognitive awareness, such as post-decision wagers, could be used to complement finding from judgement-of-confidence studies.

Considering that no study had used post-decision wagering to examine metacognitive awareness in autism, chapters three and six aimed to fill this gap in research. As previously described, post-decision wagering tasks take the same format as judgement-of-confidence tasks but participants place a wager on their answer being correct rather than rating their confidence. It has been argued that post-decision wagering is more sensitive than confidence ratings at measuring conscious awareness. Persaud et al. (2007) argues that the monetary element encourages participants to reveal all conscious awareness, overcoming issues with confidence bias (the tendency to give high/low confidence answers due to factors other than actual confidence). Others argues that this sensitivity has been overstated and that it is equal to, but not better that, confidence ratings, provided risk aversion is taken into account. (Dienes & Seth, 2010). This, therefore, raises the issue that any study that uses post-decision wagering as a measuring of metacognition when comparing group must ensure that both groups are matched for risk aversion, unless the paradigm has been designed so that participants do not stand to lose anything (no-loss paradigm). In chapters three and six of this thesis the groups were matched for risk aversion to ensure that the results were not the result of differences in risk propensity. Thus, providing that risk aversion is taken into account using both confidence ratings and post-decision wagering to explore metacognition in autism will ensure a thorough investigation.

Feeling of Knowing and Remember-Know-Guess

The final chapter in this thesis (chapter seven) aimed to examine a much under researched area of metacognition in autism, namely feeling-of-knowing. This was combined with a remember-know-guess paradigm in order to examine prospective and retrospective metacognition. Feeling-of-knowing paradigms require participants to make a judgement about the likelihood that they will perform well on a future cognitive task. Remember-know-guess paradigm on the other hand require participants to make a judgment immediately *after* completing the task.

Feeling-of-knowing tasks were one of the first tasks specifically aim at measuring metacognition (Hart, 1965) and have been used a vast number of times, a quick search of ScienceDirect returns over 50,000 results. Whilst feeling-of-knowing is a well-established and well validated measure of metacognition, remember-know-guess tasks are less straight forward. There are two main interpretations of remember-know, the dual-process theory and the signal-detection theory. The dual-process theory posits that ‘remember’ and ‘know’ responses tap into distinct states of awareness (Tulving, 1985). ‘Remember’ being associated with episodic memories which are autobiographical in nature, and ‘know’ being associated with semantic memories which involve the awareness of knowledge with no autobiographical aspect. According to the dual-process account, ‘remember’ responses reflect recollection and are viewed as categorical (one either does or doesn’t) whereas ‘know’ responses reflect familiarity and thus are viewed as continuous (high to low; Yonelinas, 2002). Given this interpretation the procedure used in chapter seven would not be tapping into metacognitive accuracy.

An alternative interpretation that has gained much support in recent years, one that would support the notion that the task utilised in chapter seven does indeed tap into metacognitive accuracy is the single process or signal-detection account (Donaldson, 1996, Dunn, 2004; Haaf et al., 2021; Rotello, Macmillan, Hicks, & Hautus, 2006; Hirshman, 1998, Inoue& Bellezza, 1998, Wixted, 2007; Wixted, & Mickes, 2010). The signal-detection account of remember-know paradigms posit that remember-know (and guess) responses reflect the strength/confidence in one’s memory rather than qualitatively difference processes. ‘Remember’ responses reflect stronger memory traces than ‘know’ responses, and ‘know’ responses reflect stronger memory traces than ‘guess’ responses. Moreover, ‘remember’ judgments are associated with higher confidence and higher accuracy than ‘know’ judgments (Wixted & Stretch, 2004). Essentially this theory posits that ‘remember’ responses are given more stringent criteria for recall/recognition than ‘know’ responses. Thus, when a participant reports that they ‘remember’ something they are indicating that they have a stronger memory and more confidence for that piece of information over the item they responded ‘know’ to. Given this logic, remember-know-guess responses can be applied in the same way that confidence ratings can be applied in order to assess metacognitive accuracy. Thus, the response provided by the participant should correlate with their performance (e.g. more correct answers for ‘remember’ responses than ‘know’ responses and more correct answers for ‘know’ responses than guess responses). Given this interpretation, one can calculate a gamma score to give an indication of metacognitive accuracy.

The debate regarding which interpretation reigns over the other has been going on since the 1990s with proponents of the dual-process interpretation highlighting data that cannot be explained by the signal detection theory and vice versa (Besson, Ceccaldi, Didic, & Barbeau, 2012; Gardiner & Java, 1990; McElree, Dolan, & Jacoby, 1999; Schacter & Tulving, 1994; Tulving, 1985; Yonelinas, 1999). The main shortcoming of the signal-detection approach in recent years being that it cannot account for 'know' responses that are accurate and associated with high confidence. This is known at the 'butcher-on-the-bus' experience after the situation described by Mandler (1980) where a person may see someone and feel like they know them, and have confidence in their knowledge but not recall any detail. This cannot be accounted for by early signal-detection theories, however, Wixted and Mickes (2010) proposed the 'Continuous Dual-Process Model', which argues that their 'Continuous Dual-Process Model' can incorporate this data and is not necessarily at odds with the dual-process interpretation but rather incorporates the idea of recollection and familiarity whilst still accounting for the signal-strength interpretation. Overall, this account can explain a large amount of data previously believed to challenge the signal-detection account as well as accounting for data that cannot be explained by the dual-process account (for a review see Wixted & Mickes, 2010).

Multiple Statistical Comparisons

Having reviewed the main measures used within this thesis, it is clear that multiple tasks and measurements have been used to answer the same questions across each chapter. This is not an uncommon approach; researchers are constantly trying to investigate the same research questions from different angles. However, doing so raises the issue that significant results may have occurred by chance. That is the more times different tests or tasks are run the more likely it is that a significant result will occur. Thus, such an approach increases the risk of type 1 errors (falsely rejecting the null). This is particularly an issue in the chapters where multiple measures of mindreading have been used. One of the most popular ways to overcome this is the Bonferroni correction, which adjusts the *p* value based on the number of tests used. The issues with this sort of correction, however, is that it reduces the power to detect true effects and thus increases the chance of type 2 errors (falsely accepting the null). This is particularly an issue where sample sizes are small, which is the case across many of the experiments in this thesis. An alternative approach that future research should consider adopting is the Bayesian perspective (Gelman, Hill & Yajima, 2012). Instead of altering the *p* value, Bayesian multi-level modelling moves point estimates and their corresponding intervals closer together based on previous data, making estimates more reliable without creating a power issue and thus reduces the chance of type 2 errors.

Measures of Metacognition

Metacognitive accuracy can be calculated in a number of ways, the most common of these being presented in Table 30. The studies presented in this thesis used gamma correlation, a non-parametric measure of metacognitive accuracy, recommended by Nelson (1984) and most commonly used in autism research. This has the advantage over the classic Pearson's correlation (also used in metacognitive research) that it can handle data that is not normally distributed. There are however a number of other issues that must be acknowledged. First, as discussed previously, gamma is influenced by object-level performance, that is a person may appear to have better metacognitive accuracy on one task compared to another simply because they found one task easier. Secondly, it suffers potentially from the effects of metacognitive bias, meaning that it does not take into account if someone has a tendency to give high/low ratings regardless of performance (Masson & Rotello, 2009). An alternative measure that overcomes both these issues is signal detection theory meta-*d'*, this however is used for binary data, although continuous data can be separated arbitrarily into high and low (e.g. confidence rating 1 = low, confidence rating 2-5 = high, Maniscalo & Lau, 2012). Thus, no measure is perfect, and each has its strengths and weaknesses. Therefore, some have argued that multiple measures of metacognitive accuracy should be used, and if there is no discrepancy then one can be more confident in their results (Dunlosky, Muller & Thiede, 2016). Dunlosky, et al., claim this in now common practice in their laboratory, and that gamma and signal detection theory have tended to support the same qualitative conclusions. They highlight that if a discrepancy was to occur this must be reported and investigated. Thus, future research into metacognition in autism should consider using multiple measures to confirm the findings.

Table 30

Characteristic of the most commonly used measures of metacognitive accuracy

Measure	Type of data	Normal distribution assumed	Metacognitive bias	Task performance
Pearson's (Phi)	Continuous and dichotomous	Yes	Yes	Yes
Gamma (G)	Continuous and dichotomous	No	Yes	Yes
Signal Detection Theory (type-2 d)	Dichotomous	Yes	No	Yes
Type 2 Receiver Operating Characteristics	Dichotomous	No	No	Yes
Signal Detection Theory (Meta- d)	Dichotomous	No	No	No

Borderline *p* values

Borderline or marginal *p* values have been reported throughout this thesis, however there is debate concerning whether it is right to report borderline or marginal *p*-values. Historically, a *p*-value of .05 has been identified as the cut off for statistical significance. More recently, the American Statistical Association (ASA, Wasserstein, Schirm, & Lazar, 2019) has produced some guidelines regarding the use of *p*-values. It is stated that conclusions should not be based purely on a *p*-value of .05 because this is an arbitrary threshold. This has been a long-standing debate (see Cohen, 1990**)** but the ASA guidelines make it clear that researchers should not just dismiss an association or group difference just because it does not meet the $p<.05$ threshold. In fact, they go as far as to say that researchers should abandon the term ‘statistically significant’ all together and should instead just report the *p*-value. Many suggestions have been put forward to address how researchers interpret data if the *p*-value of .05 is abandoned (see Wasserstein et al., 2019), and thus in line with Betensky (2019) recommendation we chose to present p-values, in the context of effect sizes and a sample size. This is particularly important given the small sample sizes across the studies presented in this thesis. Thus, as well as reporting if a result was statistically significant based on a *p*-value of .05, in line with convention, we also highlighted *p*-values that were close to the arbitrary cut off along with effect sizes for all analysis. In addition to this we included some analysis that looked at differences in effect sizes.

Mediation Analysis

Chapters five and six reported a series of mediation analyses to examine if the relationship between autism traits and metacognition was mediated by mindreading. The rational for this was that there was an association between autism traits and metacognitive accuracy across the experiments presented is these studies, and from a one-system perspective one would predict that mindreading mediates this relationship. The results from these mediation analyses indicated that mindreading ability mediated the relationship between autism traits and prospective metacognition only (see chapter five). There was no evidence of mediation for retrospective metacognition (see chapters five and six). Whilst the analysis indicated that mindreading ability mediates the relationship between autism traits and prospective metacognition, one must be cautious in this interpretation. Mediation analysis cannot prove that mindreading is causally involved, and it cannot rule out that other (unmeasured) mediators may provide an alternative/better explanation (Fiedler, Schott & Meiser, 2011). Considering that mindreading was the only mediator put into this analysis, future research should consider testing what other variables (e.g., executive function) may mediate the relationship between autism traits and metacognitive accuracy.

Generalisability and the Challenge to Metacognitive research in Autism

Another issue to consider when drawing conclusions from the studies presented in this thesis is the representativeness of the samples. It seems that across the majority of our studies the autistic participants did not have pervasive mindreading difficulties (see Table 31). It is therefore possible that the results from the studies presented in this thesis will not generalise to autistic adults who have more significant and pervasive mindreading difficulties. This is key to the theoretical debate given that the one-system view only predicts diminished metacognitive accuracy among people with diminished mindreading.

Table 31

Significance of Between-Group Differences in Mindreading for each study (Cohen's d in brackets)

Chapter	RMIE	Animation	MASC
Chapter Three	Marginal (.61)	-	Yes (.95)
Chapter Four	No (.51)	No (.40)	Yes (.92)
Chapter Five	No (.15)	No (.30)	-
Chapter Six	No (.46)	No (.45)	-
Chapter Seven	No (.48)	Marginal (.61)	

The lack of between-group differences in mindreading between our samples highlights a wider emerging issue with autism research, namely the increasing heterogeneity of autism. Since it was first described by Kanner (1943), the concept of autism has undergone many changes in line with how autism is understood, and thus the people considered to be autistic has also changed. Evidence suggests that autism has gone from a very narrow selection of individuals with very specific difficulties to a broader spectrum of people with atypical behaviours (Mottron, 2020). In line with this, the homogeneity of people diagnosed with autism has decreased. This has been reflected in the changes in research outcomes such as the reducing effect sizes in case control studies (Rødgaard, Jensen, Vergnes, Soulières, Mottron, 2019). Thus, future research should consider the pervasiveness/severity of mindreading when examining metacognition in autism or any construct that is thought to be linked to mindreading ability. For example, research could examine metacognition among groups of differing levels of mindreading ability to investigate at what point metacognition becomes impaired, if at all.

In line with this changing landscape of autism, not all the participants that took part in the current study scored above seven on the ADOS, a score of seven or above is indicative of autism (Lord et al., 2000). Across all of the studies presented in this thesis, only 63% of autistic participants (that we had scores for) scored seven or above. This does not necessarily mean that the people that scored below seven are misdiagnosed. The sensitivity of the ADOS is far from perfect, meaning that a relatively substantial proportion of individuals who are judged by clinicians to warrant an autism diagnosis do not score above seven on ADOS (Risi et al., 2006). Conversely the specificity has also been shown to be low, with 29% of non-autistic individuals being identified as autistic. Ultimately, performance on the ADOS or any other measure doesn't over-rule the diagnostic judgement of an expert clinician. However, it does mean that the results from the current thesis cannot be generalised to autistic adults with more severe difficulties, as measured by the ADOS. Thus, future autism research should factor symptom severity into the design/analysis. This could be done by employing more stringent inclusion criteria. For example, research could combine an observational measure of autism, such as the ADOS, with an interview measure of autism, such as the Autism Diagnostic Interview-Revised (ADI-R; LeCouteur, Lord, and Rutter, 2000). Evidence indicates that using both these measures improves sensitivity and specificity (Risi et al., 2006). Limiting research to those with more pervasive difficulties, would of course decrease the generalisability of results to autistic people with less severe difficulties, however, from a clinical perspective it is these people who are most likely to require significant support from services. Therefore, from a clinical perspective there is good reason to be selective when investigating difficulties that autistic individuals may have.

One way that the studies presented in the current thesis would have been able to overcome this issue would have been to examine the group differences in metacognition only using the participants that had a score of above the seven on ADOS. However, this would have resulted in very small sample sizes with very low statistical power. For example, in the chapter with the largest number of participants over the cut off (chapter three) there would have only been 16 autistic participants. This would have given a 76% chance of detecting a large effect size, but only a 43% chance of detecting a moderate effect size (≥.50). In the other chapters the statistical power would have been even smaller. Given that similar studies have found effect sizes of ranging from 0.25 to 0.66 (see chapter two) it is highly unlikely that any of the studies would have found significant differences if the participants that scored below seven were excluded.

Sample Sizes

This brings us on to the next common issue in autism research, small sample sizes. As with the majority of autism research, the case control studies presented in this thesis had relatively small sample sizes. On average there were 24 participants in each group (ASD range: 19 to 27; TD range: 20 to 27). Power analysis using G*Power 3.1 (Faul et al. 2007) indicates that to have an 80% chance of detecting a large effect size (≥.80) each group should consist of 21 participants, to detect a moderate (≥.50) effect size each group should have 51 participants, and to detect a small effect size (≥.20) each group would need 310. Thus, the majority of the case control studies in this thesis had enough statistical power to detect large effect sizes but not moderate or small. This is something that all autism research should consider when examining group differences. Therefore, future research should ensure groups are large enough to detect the predicted effect, as well as examining effect sizes and not just statistical significance. These factors are essential to ensure key difficulties within autism are not overlooked.

Summary

Overall, there is evidence to indicate that mindreading is related to metacognitive accuracy and the speed at which meta-level judgements are made, however results from this thesis were mixed. It also highlighted that autistic adults are just as accurate, albeit slower, in their meta-level judgments as typically developing adults. There is also reason to believe that whilst metacognition is accurate, it does not necessarily mean it is intact. It remains a distinct possibility that autistic adults use different strategies to achieve the same level of accuracy as typically developing adults. To be sure of this conclusion future research should investigate the processes used by autistic adults when making meta-level judgements, specifically focusing on heuristic versus analytic processes. It should also, investigate if restricting the duration autistic adults have to make their meta-level decisions, so that it is in line with typically developing adults, reduces their accuracy. Future research should also factor in severity and pervasiveness of mindreading and autistic features (social and non-social) to ensure that we develop a broader picture of metacognition across the spectrum.

References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cognitive Development, 15(1), 1-16. https://doi.org/10.1016/S0885-2014(00)00014-9

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anderson, N. J. (2002). The role of metacognition in second language teaching and learning (Vol. 4646). Washington, DC: ERIC Clearinghouse on Languages and Linguistics.

Anderson, N. J. (2012). Metacognition: Awareness of language learning. In Psychology for language learning (pp. 169-187). Palgrave Macmillan, London.

Ashwin, C., & Brosnan, M. (2019). The dual process theory of autism. In Thinking, Reasoning, and Decision Making in Autism (pp. 13-38). Routledge.

Baas, D., Castelijns, J., Vermeulen, M., Martens, R., & Segers, M. (2015). The relation between Assessment for Learning and elementary students' cognitive and metacognitive strategy use. British Journal of Educational Psychology, 85(1), 33-46.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. Science, 329(5995), 1081–1085.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. Psychological Medicine, 25, 63–77. doi:10.1017/S0033291700028099

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. Trends in cognitive sciences, 14(3), 110-118.

Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. Journal of Neuroscience, 33(42), 16657-16665.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). The lancet, 368(9531), 210-215.

Baron-Cohen, S. (1991). Precursors to a theory of mind: Understanding attention in others. Natural theories of mind: Evolution, development and simulation of everyday mindreading, 1, 233-251.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months?: The needle, the haystack, and the CHAT. The British Journal of Psychiatry, 161(6), 839-843.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”?. Cognition, 21(1), 37-46.

Baron-Cohen, S., & Swettenham, J. (1997). Theory of mind in autism: Its relationship to executive function and central coherence. Handbook of autism and pervasive developmental disorders, 2.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (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), 241–251. https ://doi.org/10.1111/1469-7610.00715 .

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (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. https ://doi.org/10.1023/A:1005653411 471.

Bartsch, K., & Wellman, H. M. (1995). Children talk about the mind. Oxford university press.

Bennetto, L., Pennington, B. F., & Rogers, S. J. (1996). Intact and impaired memory functions in autism. Child development, 67(4), 1816-1835.

Berenguer, C., Roselló, B., Colomer, C., Baixauli, I., & Miranda, A. (2018). Children with autism and attention deficit hyperactivity disorder. Relationships between symptoms and executive function, theory of mind, and behavioral problems. Research in developmental disabilities, 83, 260-269.

Besson, G., Ceccaldi, M., Didic, M., & Barbeau, E. J. (2012). The speed of visual recognition memory. Visual Cognition, 20(10), 1131-1152.

Betensky, R. A. (2019). The p-value requires context, not a threshold. The American Statistician, 73(sup1), 115-117.

Bolton, P., MacDonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., Rutter, M. (1994). A case-control family history study of autism. Journal of Child Psychology and Psychiatry, 35, 877–900. doi:10.1111/jcpp.1994.35.issue-5

Borenstein, M., Hadges, L. V., Higgins, J. P. T., Rothstein. (2009). Introduction to Meta-analysis. John Wiley & Sons: Chichester.

Bowler, D. M. (1992). “Theory of mind” in Asperger’s syndrome Dermot M. Bowler. Journal of Child Psychology and Psychiatry, 33(5), 877–893. https ://doi.org/10.1111/j.1469-7610.1992.tb01962.x.

Bowler, D. M., Gardiner, J. M., & Gaigg, S. B. (2007). Factors affecting conscious awareness in the recollective experience of adults with Asperger’s syndrome. Consciousness and cognition, 16(1), 124-143.

Bowler, D. M., Gardiner, J. M., & Grice, S. J. (2000). Episodic memory and remembering in adults with Asperger syndrome. Journal of autism and developmental disorders, 30(4), 295-304.

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.

Brosnan, M., Ashwin, C., & Lewton, M. (2017). Brief report: Intuitive and reflective reasoning in autism spectrum disorder. Journal of Autism and Developmental Disorders, 47(8), 2595-2601.

Brosnan, M., Lewton, M., & Ashwin, C. (2016). Reasoning on the autism spectrum: a dual process theory account. Journal of autism and developmental disorders, 46(6), 2115-2125. https://doi.org/10.1007/s10803-016-2742-4

Brown, A. L., & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. Child development, 1-8.

Brunsdon, V. E., & Happé, F. (2014). Exploring the ‘fractionation’ of autism at the cognitive level. Autism, 18(1), 17-30. https://doi.org/10.1177/1362361313499456

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. Cognition, 112, 337-342.

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen?. Animal Cognition, 3(4), 207-220.

Carpenter, K. L., Williams, D. M., & Nicholson, T. (2019). Putting your money where your mouth is: examining metacognition in ASD using post-decision wagering. Journal of autism and developmental disorders, 49(10), 4268-4279.

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: an integrative theory of self-knowledge. Oxford: Oxford University Press.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain, 125(8), 1839-1849.

Castelli, F., Happé, F., Frith, U., & Frith, C. (2013). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. In Social neuroscience (pp. 155-169). Psychology Press.

Chevallier, C., Parish-Morris, J., Tonge, N., Le, L., Miller, J., & Schultz, R. T. (2014). Susceptibility to the audience effect explains performance gap between children with and without autism in a theory of mind task. Journal of Experimental Psychology: General, 143(3), 972.

Cho, H.-C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? Journal of Business Research, 66(9), 1261–1266.

Chua, H. F., Gonzalez, R., Taylor, S. F., Welsh, R. C., & Liberzon, I. (2009). Decision-related loss: regret and disappointment. Neuroimage, 47(4), 2031-2040.

Chung, Y. S., Barch, D., & Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. Schizophrenia bulletin, 40(3), 602-616.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284. https ://doi.org/10.1037/1040-3590.6.4.284.

Clark, R. C. (1988). Metacognition and human performance improvement. Performance Improvement Quarterly, 1(1), 33-45.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. Cognitive development, 9(4), 377-395., (1994)

Cohen, J. (1969). Statistical power analysis for the behavioural sciences. San Diego, CA: Academic Press.

Cohen, J. (1992). Things I have learned (so far). In Annual Convention of the American Psychological Association, 98th, Aug, 1990, Boston, MA, US; Presented at the aforementioned conference.. American Psychological Association.

Coltheart, M. (1981). The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology Section A, 33(4), 497-505. https://doi.org/10.1080/146407 48108400805.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. Psychology and aging, 12(1), 50.

Constantino, J. N. (2011). The quantitative nature of autistic social impairment. Paediatric research, 69(8), 55-62. https://doi.org/10.1203/PDR.0b013e318 212ec6e.

Constantino, J. N., & Todd, R. D. (2000). Genetic structure of reciprocal social behavior. American Journal of Psychiatry, 157(12), 2043-2045. https://doi.org/10.1176/appi.ajp.157.12.2043.

Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population: A twin study. Archives of General Psychiatry, 60(5), 524–530. doi:10.1001/archpsyc.60.5.524.

Cooper, R. A., Plaisted-Grant, K. C., Baron-Cohen, S., & Simons, J.S. (2016). Reality monitoring and metamemory in adults with autism spectrum conditions. Journal of Autism and Developmental Disorders, 46(6), 2186–2198. https ://doi.org/10.1007/s10803-016-2749-x.

Corbett, B. A., Constantine, L. J., Hendren, R., Rocke, D., & Ozonoff, S. (2009). Examining executive functioning in children with autism spectrum disorder, attention deficit hyperactivity disorder and typical development. Psychiatry research, 166(2-3), 210-222.

Cosentino, S. (2014). Metacognition in Alzheimer’s disease. In The cognitive neuroscience of metacognition (pp. 389-407). Springer, Berlin, Heidelberg.

Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. Developmental Science, 18(6), 957-971. Hembacher, Lyons, & Ghetti, 2014;

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. Law and Human Behavior, 14(2), 185-191.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. Law and Human Behavior, 12(1), 41–55. https ://doi.org/10.1007/BF010 64273.

Dienes, Z., & Seth, A. (2010). Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. Consciousness and Cognition, 19(2), 674–681. https ://doi.org/10.1007/s1080 3-016-2749-x.

Dignath, C., Buettner, G., & Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively?: A meta-analysis on self-regulation training programmes. Educational Research Review, 3(2), 101-129.

Doenyas, C., Mutluer, T., Genç, E., & Balcı, F. (2019). Error monitoring in decision-making and timing is disrupted in autism spectrum disorder. Autism Research, 12(2), 239-248.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. Memory & Cognition, 24, 523–533.

Donker, A. S., De Boer, H., Kostons, D., Van Ewijk, C. D., & van der Werf, M. P. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. Educational Research Review, 11, 1-26.

Dunlosky, J., & Metcalfe, J. (2009). Metacognition. Sage Publications.

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls. In J. Dunlosky & S. K. Tauber (Eds.), The Oxford handbook of metamemory (pp. 23–37). Oxford University Press.

Dunn, J. C. (2004). Remember-know: a matter of confidence. Psychological review, 111(2), 524.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. Current directions in psychological science, 12(3), 83-87.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., et al. (2006). Introducing MASC: A movie for the assessment of social cognition. Journal of Autism and Developmental Disorders, 36(5), 623–636. https ://doi.org/10.1007/s1080 3-006-0107-0.

El Kaddouri, R., Bardi, L., De Bremaeker, D., Brass, M., & Wiersema, J. R. (2020). Measuring spontaneous mentalizing with a ball detection task: putting the attention-check hypothesis by Phillips and colleagues (2015) to the test. Psychological Research, 84(6), 1749-1757.

Elsabbagh, Mayada, Gauri Divan, Yun-Joo Koh, Young Shin Kim, Shuaib Kauchali, Carlos Marcín, Cecilia Montiel-Nava et al. "Global prevalence of autism and other pervasive developmental disorders." Autism research 5, no. 3 (2012): 160-179.

Evans, J. S. B., & Frankish, K. E. (2009). In two minds: Dual processes and beyond. Oxford: Oxford University Press.

Farrant, A., Blades, M., & Boucher, J. (1999b). Recall readiness in children with autism. Journal of Autism and Developmental Disorders, 29(5), 359-366.

Farrant, A., Boucher, J., & Blades, M. (1999a). Metamemory in children with autism. Child Development, 107-131., Blades, & Boucher, 1999a

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175–191. https ://doi.org/10.3758/BF031 93146.

Fernández-Abascal, E. G., Cabello, R., Fernández-Berrocal, P., & Baron-Cohen, S. (2013). Test-retest reliability of the ‘Reading the Mind in the Eyes’ test: A one-year follow-up study. Molecular Autism, 4(1), 33.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. Consciousness and cognition, 9(2), 288-307., Baird, and Posner, 2000.

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. Journal of Experimental Social Psychology, 47(6), 1231-1236.

Fisher, N., Happé, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. Journal of child Psychology and Psychiatry, 46(4), 409-419.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American psychologist, 34(10), 906., 1979

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1594), 1338-1349.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. Frontiers in human neuroscience, 8, 443.

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J. C. (2016). Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. Neuroscience of Consciousness, 2016(1), niw018.

Folstein, S., & Rutter, M. (1977). Infantile autism: A genetic study of 21 twin pairs. Journal of Child Psychology and Psychiatry, 18, 297–321. doi:10.1111/j.1469-7610.1977.tb00443

Fox, S. G., & Walters, H. A. (1986). The impact of general versus specific expert testimonyand eyewitness confidence upon mock juror judgment. Law and Human Behavior, 10(3), 215-228.

Frith, U. (2013). Autism and dyslexia: A glance over 25 years of research. Perspectives on Psychological Science, 8(6), 670–672. https ://doi.org/10.1177/17456 91613 50745 7.

Frith, U., Happé, F., & Siddons, F. (1994). Autism and theory of mind in everyday life. Social Development.

Frye D, Zelazo PD, Palfai T. 1995. Theory of mind and rule-based reasoning. Cognitive development 10: 483–527

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeter, K. (2003). Enhancing third-grade student' mathematical problem solving with self-regulated learning strategies. Journal of educational psychology, 95(2), 306.

Furlano, R., & Kelley, E. A. (2019). Do Children with Autism Spectrum Disorder Understand Their Academic Competencies?. Journal of autism and developmental disorders, 50(9), 3101-3113.

Furlano, R., Kelley, E. A., Hall, L., & Wilson, D. E. (2015). Self-perception of competencies in adolescents with autism spectrum disorders. Autism Research, 8(6), 761-770.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword judgments. Memory & Cognition, 18, 23–30

Gardiner, John M., Cristina Ramponi, and Alan Richardson-Klavehn. "Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses." Memory 10, no. 2 (2002): 83-98.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. Journal of research on educational effectiveness, 5(2), 189-211.

Goldman, A. I. (2006). Simulating minds: The philosophy, psychology, and neuroscience of mindreading. Oxford University Press

Gomez, C. R., & Baird, S. (2005). Identifying early indicators for autism in self-regulation difficulties. Focus on Autism and Other Developmental Disabilities, 20(2), 106–116. https ://doi.org/10.1177/10883 57605 02000 20101 .

Gopnik, A. (1990). Developing the idea of intentionality: Children's theories of mind. Canadian Journal of Philosophy, 20(1), 89-113.

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. Proceedings of the National Academy of Sciences, 113(13), 3492-3496.

Grainger, C., Williams, D. M., & Lind, S. E. (2014). Metacognition, metamemory, and mindreading in high-functioning adults with autism spectrum disorder. Journal of Abnormal Psychology, 123(3), 650.

Grainger, C., Williams, D. M., & Lind, S. E. (2016). Metacognitive monitoring and control processes in children with autism spectrum disorder: Diminished judgement of confidence accuracy. Consciousness and Cognition, 42, 65–74. https ://doi.org/10.1016/j.conco g.2016.03.003.

Grainger, C., Williams, D. M., & Lind, S. E. (2016a). Metacognitive monitoring and control processes in children with autism spectrum disorder: Diminished judgement of confidence accuracy. Consciousness and Cognition, 42, 65-74.

Grainger, C., Williams, D. M., & Lind, S. E. (2016b). Judgment of learning accuracy in high-functioning adolescents and adults with autism spectrum disorder. Journal of Autism and Developmental Disorders, 46(11), 3570-3582.

Haaf, J. M., Rhodes, S., Naveh-Benjamin, M., Sun, T., Snyder, H. K., & Rouder, J. N. (2021). Revisiting the remember–know task: Replications of Gardiner and Java (1990). Memory & Cognition, 49(1), 46-66.

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. A. Bjork (Eds.), Handbook of metamemory and memory (pp. 429–455). New York: Taylor & Francis.

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., & Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. Archives of General Psychiatry, 68, 1095–1102. doi:10.1001/archgenpsychiatry.2011.76

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. Journal of autism and Developmental disorders, 24(2), 129-154.

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. Child development, 66(3), 843-855.

Happé, F., & Frith, U. (1995). Theory of mind in autism. In Learning and cognition in autism (pp. 177-197). Springer, Boston, MA.

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. Journal of autism and developmental disorders, 36(1), 5-25.

Happé, F., & Ronald, A. (2008). The 'fractionable autism triad': a review of evidence from behavioural, genetic, cognitive and neural research. Neuropsychology review, 18(4), 287-304.

Hart, J. T. (1967). Memory and the memory-monitoring process. Journal of verbal learning and verbal behavior, 6(5), 685-691.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. Journal of educational psychology, 56(4), 208.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement?. Psychonomic bulletin & review, 19(1), 126-134.

Hayes, A. F. (2021). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford publications.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A naturalistic measure of social understanding in autism. Journal of autism and developmental disorders, 30(3), 225-236.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. journal of Educational Statistics, 6(2), 107-128.

Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. Psychological science, 25(9), 1768-1776.

Hendricks, D. (2010). Employment and adults with autism spectrum disorders: Challenges and strategies for success. Journal of Vocational Rehabilitation, 32(2), 125–134. https ://doi.org/10.3233/JVR-2010-0502.

Hermelin, B., & O'Connor, N. (1985). Logico-affective states and nonverbal language. In E. Schopler & G. Mesibov (Eds.), Communication problems in autism (pp. 283–310). New York: Plenum Press.

Hill, E. L. (2004). Executive dysfunction in autism. Trends in cognitive sciences, 8(1), 26-32.

Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. Psychological Science, 9, 61–65.

Hobson, R. P. (2014). The coherence of autism. Autism, 18(1), 6-16.

Hurlburt, R. T., Happe, F., & Frith, U. (1994). Sampling the form of inner experience in three adults with Asperger syndrome. Psychological medicine, 24(2), 385-395.

Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. Memory & Cognition, 26, 299–308.

Inquisit 4 [Computer software]. (2015). Retrieved from https://www.millisecond.com.

Jahromi, L. B., Bryce, C. I., & Swanson, J. (2013). The importance of self-regulation for the school and peer engagement of children with high-functioning autism. Research in Autism Spectrum Disorders, 7(2), 235–246.

Jones, C. R., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J., Tregay,J., et al. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. Autism Research, 11(1), 95–109.

Kawamura, Y., Takahashi, O., & Ishii, T. (2008). Reevaluating the incidence of pervasive developmental disorders: impact of elevated rates of detection through implementation of an integrated system of screening in Toyota, Japan. Psychiatry and clinical neurosciences, 62(2), 152-159.

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. Neuropsychology review, 18(4), 320-338.

Kim, Y. S., Leventhal, B. L., Koh, Y. J., Fombonne, E., Laska, E., Lim, E. C., ... & Grinker, R. R. (2011). Prevalence of autism spectrum disorders in a total population sample. American Journal of Psychiatry, 168(9), 904-912.

Knight, P. T., & Yorke, M. (2002). Employability through the curriculum. Tertiary education and management, 8(4), 261-276.

Koriat, A. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. Metacognition: Knowing about knowing, 115-135.

Koriat, A., & Helstrup, T. (2007). Metacognitive aspects of memory. In Everyday memory (pp. 263-286). Psychology Press.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings-of-knowing . Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(1), 34.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. Science, 330(6012), 1830-1834.

Kreutzer, M. A., Leonard, C., Flavell, J. H., & Hagen, J. W. (1975). An interview study of children's knowledge about memory. Monographs of the society for research in child development, 1-60

Kruskal, W. H., & Goodman, L. (1954). Measures of association for cross classifications. Journal of the American Statistical Association, 49(268), 732–764.

Kučera, H., & Francis, W. N. (1967). Computational analysis of present-day American English. Dartmouth.

Landry, O., & Chouinard, P. A. (2016). Why we should study the broader autism phenotype in typically developing populations. Journal of Cognition and Development, 17(4), 584-595.

Langdell, T. (1978). Recognition of faces: An approach to the study of autism. Journal of child psychology and psychiatry, 19(3), 255-268.

Le Couteur, A., Bailey, A., Goode, S., Pickles, A., Gottesman, I., Robertson, S., & Rutter, M. (1996). A broader phenotype of autism: The clinical spectrum in twins. Journal of Child Psychology and Psychiatry, 37, 785–801. doi:10.1111/j.1469-7610.1996.tb01475.x

LeCouteur, A., Lord, C., & Rutter, M. (2003). The Autism Diagnostic Interview – Revised (ADI-R). Los Angeles, CA: Western Psychological Services.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). Journal of Experimental Psychology: Applied, 8(2), 75. https ://doi.org/10.1037/1076-898X.8.2.75.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". Psychological review, 94(4), 412.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. Developmental science, 1(2), 247-253.

Lind, S. E., Williams, D. M., Bowler, D. M., & Peel, A. (2014). Episodic memory and episodic future thinking impairments in high-functioning autism spectrum disorder: An underlying difficulty with scene construction or self-projection. Neuropsychology, 28, 55– 67. doi:10.1037/neu0000005

Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. Neuroscience and Biobehavioral Reviews, 80, 729–742.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. Journal of experimental psychology: Human learning and memory, 4(1), 19.

Loftus, E., & Ketcham, K. (1991). Witness for the defense: The accused, the eyewitness, and the expert who puts memory on trial. Macmillan.

López, B., Donnelly, N., Hadwin, J., & Leekam, S. (2004). Face processing in high-functioning adolescents with autism: Evidence for weak central coherence. Visual Cognition, 11(6), 673-688.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders, 30(3), 205–223. https ://doi.org/10.1023/A:10055 92401 947.

Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. Child development, 84(2), 726-736.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. Psychological Review, 87, 252–271.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and cognition, 21(1), 422-430.

Maras, K., Gamble, T., & Brosnan, M. (2019). Supporting metacognitive monitoring in mathematics learning for young people with autism spectrum disorder: a classroom-based study. Autism, 23(1), 60-70.

Maras, K., Norris, J. E., & Brewer, N. (2020). Metacognitive monitoring and control of eyewitness memory reports in autism. Autism Research, 13(11), 2017-2029.

Markessini, J. (1996). Executive Leadership in a Changing World Order: Requisite Cognitive Skills. The First Literature Review.

Marshall-Mies, J. C., Fleishman, E. A., Martin, J. A., Zaccaro, S. J., Baughman, W. A., & McGee, M. L. (2000). Development and evaluation of cognitive and metacognitive measures for predicting leadership potential. The Leadership Quarterly, 11(1), 135-153.

Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(2), 509.

McCartney, K., & Burchinal, M. R. (2006). Best practices in quantitative methods for developmentalists. Monographs of the Society for Research in Child Development, 71(3), i-145.

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: a time course analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(3), 563.

McMahon, C. M., Henderson, H. A., Newell, L., Jaime, M., & Mundy,P. (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. https ://doi.org/10.1007/s1080 3-015-2630-3.

Mervis, C. B., & Klein-Tasman, B. P. (2004). Methodological issues in group-matching designs: α levels for control variable comparisons and measurement characteristics of control and target variables. Journal of Autism and Developmental Disorders, 34(1), 7-17.

Metcalfe, J. (2009). Metacognitive judgments and control of study. Current Directions in Psychological Science, 18(3), 159–163.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. Psychonomic Bulletin & Review, 15(1), 174–179.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(4), 851.

Meyer, B. J., Gardiner, J. M., & Bowler, D. M. (2014). Directed forgetting in high-functioning adults with autism spectrum disorders. Journal of Autism and Developmental Disorders, 44(10), 2514-2524.

Mikulecky, L., & Ehlinger, J. (1986). The influence of metacognitive aspects of literacy on job performance of electronics technicians. Journal of Reading Behavior, 18(1), 41-62.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive psychology, 41(1), 49-100.

Moely, B. E., Hart, S. S., Leal, L., Santulli, K. A., Rao, N., Johnson, T., & Hamilton, L. B. (1992). The teacher's role in facilitating memory and study strategy development in the elementary school classroom. Child Development, 63(3), 653-672.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine, 6(7), e1000097.

Mottron, L. (2021). A radical change in our autism research strategy is needed: Back to prototypes. Autism Research.

Mottron, L., & Bzdok, D. (2020). Autism spectrum heterogeneity: fact or artifact?. Molecular Psychiatry, 25(12), 3178-3185.

Mumford, M. D., Baughman, W. A., Supinski, E. P., Costanza, D. P., & Threlfall, K. V. (1993). Cognitive and metacognitive skill development: Alternative measures for predicting leadership potential. Management Research Institute, Bethesda, MD, US Army Research Institute for Behavioral and Social Sciences Rep. No. SBIR A92-154.

Nebel-Schwalm, M., & Worley, J. (2014). Other disorders frequently comorbid with autism. In Handbook of autism and anxiety (pp. 47-60). Springer, Cham.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin, 95(1), 109. https ://doi.org/10.1037/0033-2909.95.1.109.

Nelson, T. O., & Narens, L. (1990). Metamemory a theoretical framework and new findings. The Psychology of Learning and Motivation, 26, 125–174.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment Recall and Monitoring (PRAM). Psychological Methods, 9(1), 53. https ://doi.org/10.1037/1082-989X.9.1.53.

Nichols, S., & Stich, S. (2003). How to read your own mind: A cognitive theory of self-consciousness. Consciousness: New philosophical Essays, 157–200.

Nicholson, T., Williams, D. M., Grainger, C., Christensen, J., Calvo-Merino, B., & Gaigg, S. (2018). Interoceptive impairments do not lie at the heart of autism or alexithymia. Journal of Abnormal Psychology, 127, 612–622.

Nicholson, T., Williams, D. M., Grainger, C., Lind, S. E., & Carruthers, P. (2019). Relationships between implicit and explicit uncertainty monitoring and mindreading: Evidence from autism spectrum disorder. Consciousness and Cognition, 70, 11-24.

Nicholson, T., Williams, D. M., Lind, S. E., Grainger, C., & Carruthers, P. (2020). Linking metacognition and mindreading: Evidence from autism and dual-task investigations. Journal of Experimental Psychology: General.

Nonose, K., Kanno, T., & Furuta, K. (2014). Effects of metacognition in cooperation on team behaviors. Cognition, technology & work, 16(3), 349-358.

Oakley, B. F. M., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. Journal of Abnormal Psychology, 125, 818–823.

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. Metacognition and Learning, 13(2), 179–212. https ://doi.org/10.1007/s1140 9-018-9183-8.

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. Frontiers in psychology, 6, 1503.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. science, 308(5719), 255-258.

Ozonoff, S. (1995). Reliability and validity of the Wisconsin card sorting test in studies of autism. Neuropsychology, 9(4), 491.

Ozonoff, S., & Jensen, J. (1999). Brief report: Specific executive function profiles in three neurodevelopmental disorders. Journal of autism and developmental disorders, 29(2), 171-177.

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. Development and psychopathology, 6(3), 415-431.

Ozonoff, S., & Strayer, D. L. (1997). Inhibitory function in nonretarded children with autism. Journal of autism and developmental disorders, 27(1), 59-77.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. Journal of child Psychology and Psychiatry, 32(7), 1081-1105.

Ozonoff, S., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive function abilities in autism and Tourette syndrome: An information processing approach. Journal of child Psychology and Psychiatry, 35(6), 1015-1032.

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. Consciousness and cognition, 28, 151-160.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. Journal of Neuroscience Methods, 162(1–2), 8–13. https ://doi. org/10.1016/j.jneum eth.2006.11.017.

Perner, J. (1991). Understanding the representational mind. The MIT Press.

Perner, J. (2014). Development of metacognifion. Self-awareness, 14.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. Social neuroscience, 1(3-4), 245-258.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. Trends in cognitive sciences, 16(10), 519-525.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep?. Science, 308(5719), 214-216.

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that…” attribution of second-order beliefs by 5-to 10-year-old children. Journal of experimental child psychology, 39(3), 437-471.

Perry, J., Lundie, D., & Golder, G. (2019). Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools?. Educational Review, 71(4), 483-500.

Perry, V., Albeg, L., & Tung, C. (2012). Meta-analysis of single-case design research on self-regulatory interventions for academic performance. Journal of behavioral education, 21(3), 217-229.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. Nature Neuroscience, 10(2), 257. https ://doi.org/10.1038/nn184 0.

Phillips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and in autism. British Journal of Developmental Psychology, 16(3), 337-348.

Phillips, J., Ong, D. C., Surtees, A. D. R., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A second look at automatic theory of mind: Reconsidering Kovács, Téglás, and Endress (2010). Psychological Science,26(9), 1353–1367. https://doi.org/10.1177/0956797614558717.

Pickles, A., Starr, E., Kazak, S. İ. B. E. L., Bolton, P., Papanikolaou, K., Bailey, A., ... & Rutter, M. (2000). Variable expression of the autism broader phenotype: findings from extended pedigrees. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 41(4), 491-502. https://doi.org/10.1111/1469-7610.00634. et al., 2000

Pishghadam, R., & Khajavy, G. H. (2013). Intelligence and metacognition as predictors of foreign language achievement: A structural equation modeling approach. Learning and Individual Differences, 24, 176-181.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. Behavioral and brain sciences, 1(4), 515-526.

Prior, M., & Hoffmann, W. (1990). Brief report: Neuropsychological testing of autistic children through an exploration with frontal lobe tests. Journal of autism and developmental disorders, 20(4), 581-590.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., ... & Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. Journal of the American Academy of Child & Adolescent Psychiatry, 45(9), 1094-1103.

Ratcliff, R. & Starns, J. J. 2009 Modeling confidence and response time in recognition memory. Psychol. Rev. 116, 59–83. (doi:10.1037/a0014086)

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. Psychological review, 120(3), 697.

Rødgaard, E. M., Jensen, K., Vergnes, J. N., Soulières, I., & Mottron, L. (2019). Temporal changes in effect sizes of studies comparing individuals with and without autism: a meta-analysis. *JAMA psychiatry*, *76*(11), 1124-1132.

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

Ronald, A., Happé, F., Bolton, P., Butcher, L. M., Price, T. S., Wheelwright, S., ... & Plomin, R. (2006). Genetic heterogeneity between the three components of the autism spectrum: a twin study. Journal of the American Academy of Child & Adolescent Psychiatry, 45(6), 691-699. https://doi.org/10.1097/01.chi.0000215325.13058.9det al., 2006

Rotello, C. M., Macmillan, N. A., Hicks, J. L., & Hautus, M. J. (2006). Interpreting the effects of response bias on remember-know judgments using signal detection and threshold models. Memory & Cognition, 34(8), 1598-1614.

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: insights from individual differences and neuroimaging. Personality neuroscience, 1.

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. Journal of Experimental Child Psychology, 80(3), 201–224. https ://doi.org/10.1006/jecp.2001.2633.

Rumsey, J. M., Rapoport, J. L., & Sceery, W. R. (1985). Autistic children as adults: Psychiatric, social, and behavioral outcomes. Journal of the American Academy of Child Psychiatry, 24(4), 465-473.

Russell, J., & Hill, E. L. (2001). Action-monitoring and intention reporting in children with autism. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42(3), 317-328.

Şahİn, S. M., & Kendir, F. (2013). The effect of using metacognitive strategies for solving geometry problems on students achievement and attitude. Educational Research and Reviews, 8(19), 1777-1792.

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. Nature neuroscience, 7(5), 499-500.

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? Consciousness and Cognition, 19(4), 1069–1078. https :// doi.org/10.1016/j.concog.2009.12.013.

Sawyer, A. C., Williamson, P., & 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.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. Psychological science, 17(8), 692-699.

Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: Evidence from patients with lesions to frontal cortex. Neuropsychologia, 42(7), 957-966.

Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. Educational Psychology Review, 20(4), 463–467.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neuroscience & Biobehavioral Reviews, 42, 9-34.

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. Journal of child Psychology and Psychiatry, 24(4), 613-620.

Shattuck, P. T., Seltzer, M. M., Greenberg, J. S., Orsmond, G. I., Bolt, D., Kring, S., et al. (2007). Change in autism symptoms and maladaptive behaviors in adolescents and adults with an autism spectrum disorder. Journal of Autism and Developmental Disorders, 37(9), 1735–1747.https ://doi.org/10.1007/s1080 3-006-0307-7.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. Journal of the American Academy of Child & Adolescent Psychiatry, 47(8), 921-929.

Souchay, C., Wojcik, D. Z., Williams, H. L., Crathern, S., & Clarke, P. (2012). Recollection in adolescents with Autism Spectrum Disorder. Cortex A Journal Devoted to the Study of the Nervous System and Behavior. doi:10.1016/j.cortex.2012.07.011.

Steel, M. J. G., Gorman, R., & Flexman, J. E. (1984). Neuropsychiatric testing in an autistic mathematical idiot-savant: Evidence for nonverbal abstract capacity. Journal of the American Academy of Child Psychiatry, 23(6), 704-707.

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: a free and simple tool for meta-analysis. Research synthesis methods, 8(4), 537-553.

Tobias, S., & Everson, H. T. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education. In Handbook of metacognition in education (pp. 119-140). Routledge.

Tulving, E. (1985). Memory and consciousness. Canadian Psychology/Psychologie canadienne, 26(1), 1.

Van Laar, E., Van Deursen, A. J., Van Dijk, J. A., & De Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. Computers in human behavior, 72, 577-588.

Veenman, M. V., & Beishuizen, J. J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraint. Learning and Instruction, 14(6), 621-640.

Veenman, M. V., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. Learning and Individual Differences, 15(2), 159-176. https://doi.org/10.1016/j.lindif.2004.12.001.

Veenman, M. V., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills in early adolescence. Instructional Science, 33(3), 193-211.

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. Metacognition and learning, 1(1), 3-14.

Veenman, M. V., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. Learning and instruction, 14(1), 89-109.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p< 0.05”. The American Statistician, 73(sup1), 1-19.

Wechsler, D. (1999). Wechsler abbreviated scale of intelligence. New York: The Psychological Corporation: Harcourt Brace & Company.

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., ... & Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. Consciousness and cognition, 22(1), 264-271.

Wellman, H. M., & Johnson, C. N. (1979). Understanding of mental processes: a developmental study of" remember" and" forget". Child Development, 79-88.

Wellman, H., Cross, D. & Watson, J. (2001) ‘Meta-Analysis of Theory of Mind Development: The Truth about False Belief’, Child Development 72: 655–684.

White, S. J., & Saldaña, D. (2011). Performance of children with autism on the Embedded Figures Test: A closer look at a popular task. Journal of autism and developmental disorders, 41(11), 1565-1572

White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on “open-ended” executive function tests in autism. Autism Research, 2(3), 138-147.

Wilkinson, D. A., Best, C. A., Minshew, N. J., & Strauss, M. S. (2010). Memory awareness for faces in individuals with autism. Journal of Autism and Developmental Disorders, 40(11), 1371–1377. https ://doi.org/10.1007/s1080 3-010-0995-x.

Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness?. Autism, 14(5), 474-494.https ://doi.org/10.1177/13623 61310 36631 4.

Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness?. Autism, 14(5), 474-494.https ://doi.org/10.1177/13623 61310 36631 4.

Williams, D. M., & 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(6), 865-873.

Williams, D. M., Bergström, Z., & Grainger, C. (2018). Metacognitive monitoring and the hypercorrection effect in autism and the general population: Relation to autism (-like) traits and mindreading. Autism, 22(3), 259–270. https ://doi.org/10.1177/1362361316 68017 8.

Williams, D. M., Lind, S. E., & Happé, F. (2009). Metacognition may be more impaired than mindreading in autism. Behavioral and Brain Sciences, 32(2), 162–163. https ://doi.org/10.1017/S0140525X0 90008 18.

Williams, D., & Happé, F. (2010). Representing intentions in self and other: Studies of autism and typical development. Developmental science, 13(2), 307-319.

Williams, D.M., & Jarrold, C. (2013). Assessing planning and set-shifting abilities in autism: Are experimenter-administered and computerised versions of tasks equivalent? Autism Research, 6, 461-467.

Wilson, A. C. (2021). Do animated triangles reveal a marked difficulty among autistic people with reading minds?. Autism, 1362361321989152.

Wilson, C. E., Happé, F., Wheelwright, S. J., Ecker, C., Lombardo, M. V., Johnston, P., et al. (2014). The neuropsychology of male adults with high-functioning autism or Asperger syndrome. Autism Research, 7(5), 568–581.https ://doi.org/10.1002/aur.1394

Wilson, D. B., Ph.D. (2020). Practical Meta-Analysis Effect Size Calculator [Online calculator]. Retrieved April 15, 2020, from https://campbellcollaboration.org/research-resources/effect-size-calculator.html, 2020

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition, 13(1), 103-128.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. Psychological review, 114(1), 152.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. Psychological review, 117(4), 1025.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. Psychonomic Bulletin & Review, 11, 616–641.

Wojcik, D. Z., Allen, R. J., Brown, C., & Souchay, C. (2011). Memory for actions in autism spectrum disorder. Memory, 19(6), 549–558. https ://doi.org/10.1080/09658 211.2011.59050 6.

Wojcik, D. Z., Moulin, C. J., & Souchay, C. (2013). Metamemory in children with autism: Exploring “feeling-of-knowing” in episodic and semantic memory. Neuropsychology, 27(1), 19.

Wojcik, D. Z., Waterman, A. H., Lestié, C., Moulin, C. J., & Souchay, C. (2014). Metacognitive judgments-of-learning in adolescents with autism spectrum disorder. Autism, 18(4), 393-408.

Woolley, J. D., & Wellman, H. M. (1990). Young children's understanding of realities, nonrealities, and appearances. Child development, 61(4), 946-961.

World Health Organisation (2021, September, 11). Autism Spectrum Disorder. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders

World Health Organization. (2018). International classification of diseases for mortality and morbidity statistics (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/en

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1594), 1310-1321. https://doi.org/10.1098/rstb.2011.0416.

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

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characterstics. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(6), 1415.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language, 46, 441–517.

Zaccaro, S. J., Mumford, M. D., Marks, M. A., Connelly, M. S., Threlfall, K. V., Gilbert, J., & Fleishman, E. A. (1997). Cognitive and temperament determinants of Army leadership. Management Research Institute, Bethesda, MD.

Zalla, T., Miele, D., Leboyer, M., & Metcalfe, J. (2015). Metacognition of agency and theory of mind in adults with high functioning autism. Consciousness and cognition, 31, 126-138.

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. Psychological Science, 7(5), 294-300.

Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: Current and future directions. Studies in Science education, 49(2), 121-169.

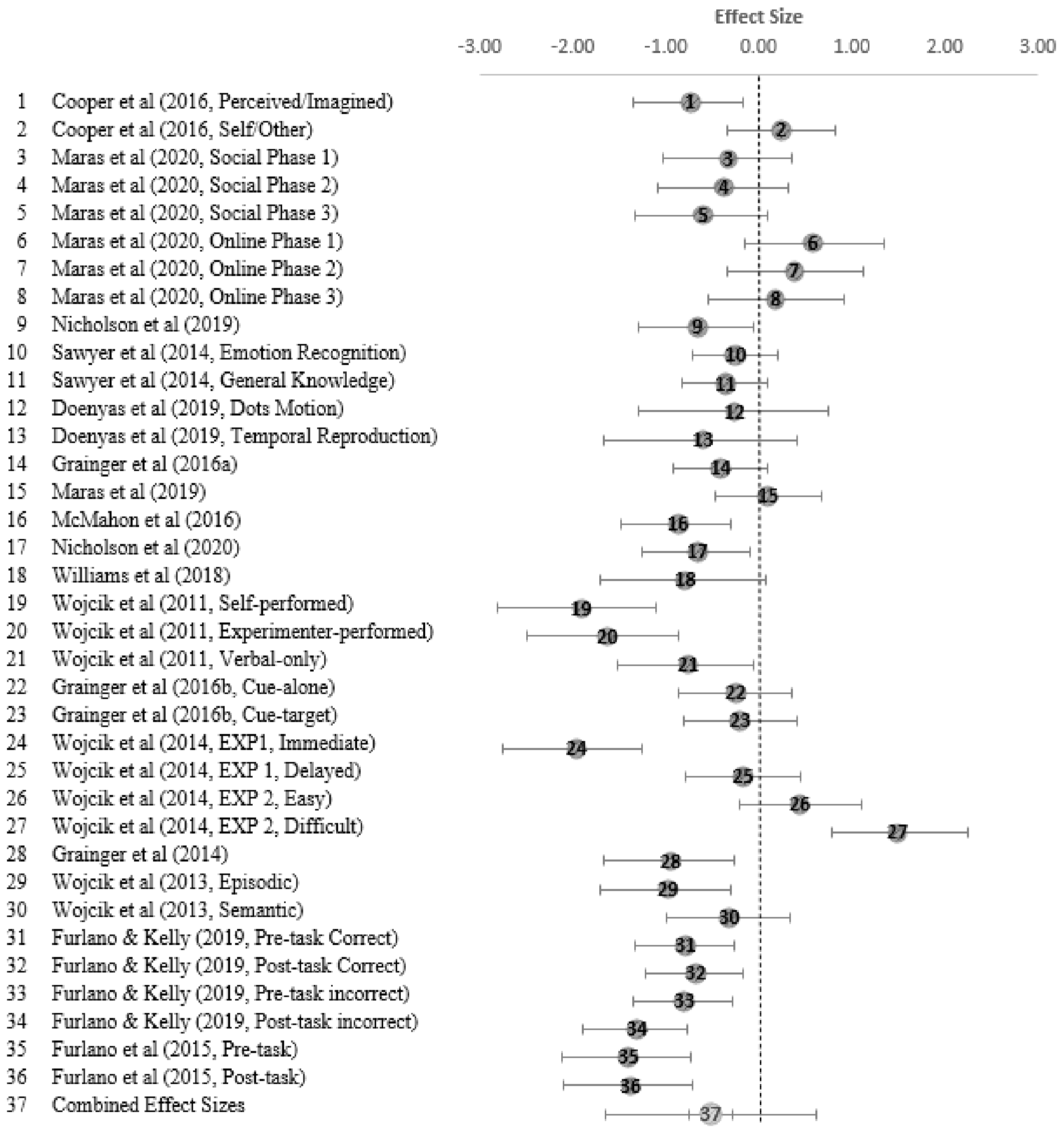
Appendices

Appendix 1: Results from Meta-analysis with Unaveraged Effect Sizes

When each individual effect size was entered without averaging the weighted effect size for the between-group difference in meta-monitoring ability -0.52 (SE 0.11, 95%CI -0.75 to -0.29) and statistically significant, z=-4.53, *p*< .001. As with the averaged effect sizes this suggests a moderate, although increased, impairment of metacognitive accuracy among the autism groups in comparison to the typically developing groups. However, as with the averaged effect sizes the homogeneity test was significant (*Q*=141.17, *p*=<.001), The effect sizes and accompanying confidence intervals are presented in Figure 3.

Figure 4

Forest Plot and Confidence Intervals from Unaveraged Effect Size Meta-Analysis.



Appendix 2: Words and their Kucera-Francis written frequency included in each trial of the Semantic Wagering Task

Main Word	KF	Synonym	KF	Lure	KF	Total KF
Propitious	2	Auspicious	1	Abridgment	1	4
Fastidious	3	Meticulous	1	Mendacious	1	5
Fecund	1	Prolific	2	Palpable	2	5
Repose	2	Quiescent	2	Rancorous	2	6
Irritable	5	Fractious	1	Facetious	1	7
Judicious	1	Discreet	3	Disgrace	3	7
Friable	2	Brittle	3	Bloated	3	8
Insolent	2	Impudent	3	Infallible	3	8
Mendacious	1	Deceptive	4	Delicious	4	9
Verdant	1	Unripe	7	Unsafe	1	9
Loquacious	2	Talkative	4	Thrilling	4	10
Sickening	2	Repulsive	4	Rehearsal	4	10
Recalcitrant	2	Resistant	5	Reverence	5	12
Fake	10	Spurious	2	Spavined	1	13
Conscientious	10	Diligent	2	Decadent	2	14
Guileless	1	Naïve	7	Slick	7	15
Stubborn	12	Recalcitrant	2	Recapitulate	1	15
Ineluctable	1	Inescapable	8	Inclination	8	17
Abandon	17	Forsake	1	Fallacy	1	19
Collect	16	Amass	2	Adore	2	20

Profuse	2	Abundant	9	Anecdote	9	20
Profess	5	Confess	11	Chronic	11	27
Loud	20	Noisy	6	Needy	6	32
Select	23	Designate	5	Developer	5	33
Calamitous	2	Disastrous	16	Compatible	16	34
Luminous	12	Shining	21	Seclude	1	34
Adroit	1	Clever	17	Absurd	17	35
Inveterate	3	Persistent	16	Obligation	16	35
Supercilious	1	Indifference	17	Productivity	17	35
Pernicious	1	Deadly	19	Exceed	19	39
Brave	24	Deduct	12	Daring	12	48
Weak	32	Frail	8	Flair	8	48
Describe	41	Portray	6	Prelude	5	52
Insidious	2	Subtle	25	Terror	25	52
Balmy	2	Gentle	27	Excuse	27	56
Powerful	63	Pungent	4	Parable	3	70
Comely	1	Attractive	39	Inevitably	38	78
Authentic	20	Genuine	34	Gesture	32	86
Gigantic	10	Enormous	37	Estimate	39	86
Famous	89	Exalted	7	Eminent	9	105
Remarkable	47	Extraordinary	31	Entertainment	29	107
Fast	78	Homeric	15	Hastily	15	108
Awful	17	Terrible	45	Tendency	49	111
Hit	115	Collision	7	Consensus	7	129

Trenchant	4	Sharp	72	Broke	72	148
Trouble	134	Plight	7	Postal	7	148
Tell	268	Inform	7	Invent	7	282
Limpid	1	Simple	161	Higher	160	322
Pretty	107	Beautiful	127	Proper	95	329
Take	611	Glow	16	Grab	16	643

Appendix 3: Words and their Kucera-Francis written frequency (KF) included in the Feeling-of-knowing/Remember Know Task

Table??

Cue	KF	Target	KF	Lure 1	KF	Lure 2	KF	Incorrect target	KF
Spoon	6	Towel	6	Snack	6	Pants	9	Bacon	10
Duck	9	Coin	10	Lamb	7	Doll	10	Lamp	18
Bird	31	Knee	35	Fish	35	Rice	33	Bowl	23
Coat	43	Skin	47	Meat	45	Dirt	43	Coin	10
Tooth	20	Bowl	23	Boot	13	Wave	46	Skin	47
Bell	18	Lamp	18	Lion	17	Chin	27	Knee	35
Sugar	34	Fruit	35	Stick	39	Paint	37	Tiger	7
Glove	9	Tiger	7	Thumb	10	Couch	12	Fruit	35
Ring	47	Lake	54	Salt	46	Mail	47	Seed	41
Belt	29	Gift	33	Sand	28	Pond	25	Lake	54
Bush	14	Leaf	12	Bull	14	Rope	15	Cart	5

Mouse	10	Bacon	10	Elbow	10	Berry	9	Drain	18
Gate	37	Seed	41	Cash	36	Tool	40	Gift	33
Desk	65	Flat	41	Fort	55	Soil	54	Pole	18
Skirt	21	Drain	18	Cream	20	Flame	17	Chick	3
Jail	21	Pole	18	Lock	23	Chip	17	Flat	41
Snail	1	Chick	3	Crust	1	Jewel	1	Towel	6
Vest	4	Cart	5	Crib	5	Toad	4	Leaf	12
Bulb	7	Shed	11	Cave	9	Calf	11	Curb	13
Silk	12	Curb	13	Tile	16	Fuel	17	Shed	11
Bolt	6	Bean	5	Cork	9	Clog	2	Drum	11
Flea	2	Drum	11	Cage	9	Frog	1	Bean	5
Zebra	1	Tunic	1	Prune	1	Dingo	1	Venom	2
Bunny	1	Chive	1	Bison	1	Camel	1	Tunic	1
Basil	1	Venom	2	Vault	2	Puppy	2	Chive	1

Curry	2	Broom	2	Thorne	3	Valve	3	Peach	3
Skull	3	Peach	3	Grape	3	Jelly	3	Canal	3
Dummy	3	Crumb	3	Clown	3	Cloak	3	Gravy	4
Chalk	3	Canal	3	Syrup	4	Yacht	4	Broom	2
Spice	4	Gravy	4	Fairy	4	Scarf	4	Crumb	3
Photo	5	Wrist	5	Witch	5	Swamp	5	Badge	5
Olive	5	Flask	5	Dingy	5	Eagle	5	Trunk	8
Canon	5	Badge	5	Spine	6	Buggy	6	Sword	7
Tiger	7	Sword	7	Spear	7	Canoe	7	Flask	5
Apron	7	Trunk	8	Flour	8	Ankle	8	Wrist	5
Pearl	9	Glove	9	Token	10	Wheat	9	Cliff	11
Spade	10	Razor	15	Cigar	10	Organ	12	Arrow	14
Juice	11	Cliff	11	Tower	13	Shelf	12	Brick	18
Purse	14	Arrow	14	Stove	15	Candy	16	Glove	9

Steam	17	Brick	18	Lemon	18	Angel	18	Razor	15
Toast	19	Crown	19	Slide	20	Screw	21	Swing	24
Slate	20	Clock	20	Shell	22	Plate	22	Crown	19
Cabin	23	Swing	24	Coach	24	Phone	25	Honey	25
Penny	25	Honey	25	Adult	25	Shirt	27	Pound	28
Prize	28	Pound	28	Movie	29	Cloud	28	Clock	20
Brush	44	Ocean	34	Grass	53	Piano	38	Blind	47
Cloth	43	Bread	41	Snake	44	Brain	45	Truck	57
Chain	50	Blind	47	Wagon	55	Chest	53	Block	66
Stone	58	Truck	57	Metal	61	Beach	61	Bread	41
Chair	66	Block	66	Kinfe	76	Sweet	70	Ocean	34
Average KF	18.6		18.7		19.6		19.7		18.7